

VŠB – Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra kybernetiky a biomedicínského inženýrství

Metody shlukové analýzy pro biomedicínské signály

Methods of Clustering Analysis for Biomedical Signals

2019

Jiří Škandera

Zadání bakalářské práce

Student:

Jiří Škandera

Studijní program:

B2649 Elektrotechnika

Studijní obor:

3901R039 Biomedicínský technik

Téma:

Metody shlukové analýzy pro biomedicínské signály
Methods of Clustering Analysis for Biomedical Signals

Jazyk vypracování:

čeština

Zásady pro vypracování:

1. Nastudování základních principů klasifikace biomedicínských dat.
2. Rešerše metod shlukové analýzy.
3. Rešerše objektivizačních parametrů pro hodnocení efektivity metod shlukové analýzy.
4. Nastudování principů metody s učitelem a bez učitele.
5. Design a realizace vybraných konvenčních metod shlukové analýzy.
6. Design a realizace vybraných nekonvenčních metod shlukové analýzy.
7. Aplikace metod shlukové analýzy pro vybrané 1D a 2D biomedicínské signály.
8. Testování a evaluace metod shlukové analýzy na základě vybraných parametrů.
9. Zhodnocení dosažených výsledků.

Seznam doporučené odborné literatury:

- [1] EVERITT, Brian. *Cluster analysis*. 5th ed. Chichester, West Sussex, U.K.: Wiley, 2011. Wiley series in probability and statistics. ISBN 978-0-470-74991-3.
- [2] EVERITT, Brian, Sabine LANDAU a Morven LEESE. *Cluster analysis*. 4th ed. New York: Oxford University Press, 2001. ISBN 0-340-76119-9.
- [3] ABONYI, Janos a Balázs FEIL. *Cluster analysis for data mining and system identification*. Boston: Birkhäuser, c2007. ISBN 978-3-7643-7987-2.

Formální náležitosti a rozsah bakalářské práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí bakalářské práce: **Ing. Jan Kubiček, Ph.D.**

Datum zadání: 01.09.2018

Datum odevzdání: 30.04.2019

doc. Ing. Jiří Koziorek, Ph.D.
vedoucí katedry



prof. Ing. Pavel Brandštetter, CSc.
děkan fakulty

Prohlášení studenta

„Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně. Uvedl jsem všechny literární
prameny a publikace, ze kterých jsem čerpal.“

V Ostravě dne: *30.dubna 2019*



.....
podpis studenta

Poděkování

Rád bych touto cestou poděkoval mému vedoucímu práce panu Ing. Janu Kubíčkovi, Ph.D. za jeho cenné rady, připomínky a odborné konzultace při vytváření této bakalářské práce.

Abstrakt

Bakalářská práce se zabývá regionální segmentací obrazu na základě shlukové analýzy. Úkolem této práce je realizace a design vybraných nehierarchických metod k-means a fuzzy c-means. Testování bylo provedeno na reálných medicínských datech, která jsou ovlivněna umělým aditivním šumem. Součástí práce je kapitola zabývající se modelováním tkání na vybraných medicínských obrazech za účelem jejich následné extrakce a kapitola pro implementaci shlukové analýzy na 1D EKG signály s cílem dekompozice charakteristických segmentů těchto signálů. Klíčové testování proběhlo na 3 datasetech medicínských obrazů a vyhodnocení výsledku pro jednotlivé obrazy bylo stanoveno na základě korelačního koeficientu a střední kvadratické chyby. Posledním krokem bylo zhodnocení dosažených výsledků přenesených především do grafické podoby. Následovala tvorba graficko-uživatelského rozhraní pro zjednodušení testování. Výsledkem práce je dynamická evaluace nehierarchických shlukovacích technik pro tvorbu matematických modelů tkání v závislosti na intenzitě šumu.

Klíčová slova

Shluková analýza, segmentace obrazu, k-means, fuzzy c-means, učení bez učitele, MATLAB

Abstract

The bachelor thesis deals with regional image segmentation based on cluster analysis. The aim of this work is realisation and design of selected non-hierarchical methods k-means and fuzzy c-means. Testing was performed on real medical data which is affected by artificial additive noise. Part of this work is a chapter dealing with tissue modeling on selected medical images for the purpose of their subsequent extraction and chapter for implementation of cluster analysis on 1D ECG signals with the aim of decomposing characteristic segments of these signals. Main reason for testing was performed on 3 medical image datasets and the result was evaluated based on the correlation coefficient and the mean square error. The last step was to evaluate the achieved results transferred mainly to the graphic form. This was followed by the creation of a graphical-user interface to simplify testing. The result of this work is a dynamic evaluation of non-hierarchical clustering techniques for creating mathematical models of tissues depending on the intensity of noise.

Key words

Cluster analysis, image segmentation, k-means, fuzzy c-means, unsupervised learning, MATLAB

Obsah

Seznam použitých symbolů a zkratk.....	8
Seznam obrázků.....	9
Seznam tabulek.....	12
Úvod	13
1 Strojové učení.....	14
1.1 Učení s učitelem.....	14
1.1.1 Klasifikace.....	14
1.1.2 Regrese	14
1.2 Učení bez učitele.....	15
2 Podobnost objektů	16
2.1 Metriky vzdálenosti mezi dvěma obrazy s kvantitativními příznaky	16
2.1.1 Euklidovská vzdálenost	16
2.1.2 Manhattan vzdálenost.....	17
2.1.3 Minkowského vzdálenost	17
2.2 Metriky podobnosti dvou obrazů s kvantitativními příznaky	17
2.2.1 Skalární součin	17
2.2.2 Tanimotova metrika podobnosti.....	17
3 Shluková analýza	18
3.1 Hierarchické metody	18
3.1.1 Aglomerativní metody.....	19
3.1.2 Divizní metody	21
3.2 Nehierarchické metody.....	21
3.2.1 Metoda k-means	21
3.2.2 Metoda PAM	22
3.2.3 Fuzzy c-means	22
3.2.4 Metoda CLARA	23
3.2.5 Metoda CLARANS (Clustering Large Application based on randomised Search)	23
3.2.6 Bagged shlukování	24
4 Experimentální část.....	25
4.1 Načtení medicínských obrazů	25
4.2 Implementace šumu	26
4.3 Implementace algoritmu	26
4.4 Evaluační metriky	27
5 Implementace šumu.....	29
5.1 Gaussovský šum.....	29
5.2 Šum sůl a pepř.....	30
5.3 Šum speckle	31
5.4 Šum localvar	31
6 Implementace metod shlukové analýzy na dvourozměrné signály	32
6.1 Metoda k-means.....	32
6.2 Metoda fuzzy c-means	34
7 Testování metod na reálných medicínských datech	37
8 Modelování tkaní na základě shlukové analýzy.....	42
9 Implementace metod shlukové analýzy na jednorozměrné signály	43

10	Evaluace testovaných dat.....	46
10.1	Analýza pro rozdílný počet shlukovacích tříd	46
10.2	Analýza metody k-means a fuzzy c-means pro každý typ šumu	47
10.3	Analýza míry zkreslení obrazu na základě zvolení typu šumu	48
10.4	Rozdíl ovlivnění obrazu šumem v závislosti na typu datasetu	48
11	Návrh graficko-uživatelského prostředí.....	49
12	Závěr.....	52
	Zdroje	53
	Seznam Příloh	55

Seznam použitých symbolů a zkratek

1D	Jednorozměrný
2D	Dvojměrný
CLARA	Metoda shlukové analýzy (Clustering Large Applications)
CLARANS	Metoda shlukové analýzy (Clustering Large Application based on randomised Search)
CT	Výpočetní tomografie
D	Vzdálenost
EKG	Elektrokardiogram
MATLAB	Matrix laboratory
MRI	Magnetická rezonance
PAM	Metoda shlukové analýzy (Partitioning Around Medoids)
PQ	Interval EKG signálu
QRS	Komplex vln EKG signálu
RGB	Barevný model
ST	Úsek EKG signálu
d	Denzita
k	Počet shluků
mse	Střední kvadratická chyba
px	Pixel, obrazový bod
uint-8	Celočíselný datový typ
μ	Střední hodnota
σ	Rozptyl

Seznam obrázků

Obr. 1: Dělení strojového učení [2].....	14
Obr. 2: Znázornění Euklidovské vzdálenosti v prostoru [3]	16
Obr. 3: Rozdělení objektů do shluků [vlastní]	18
Obr. 4: Dendogram [vlastní]	19
Obr. 5: Metoda nejbližšího souseda [13]	19
Obr. 6: Metoda nejvzdálenějšího souseda [13]	20
Obr. 7: Metoda průměrné vazby [13]	20
Obr. 8: Wardova metoda [13]	20
Obr. 9: Centroidová metoda [14]	20
Obr. 10: Znázornění postupu tvoření shluků za pomoci divizních metod [14].....	21
Obr. 11: „Ukázka rozdělení objektů do shluků nehierarchickou metodou k-means. Výsledek je ovlivněn volbou počtu shluků. Vlevo: počet shluků $k=3$ je dobrá volba; vpravo: počet shluků $k=2$ je špatná volba“ [14].....	22
Obr. 12: Rozdělení shluků pro metodu PAM [vlastní]	22
Obr. 13: Model aplikování metody fuzzy c-means, volba počtu shluků, nastavení míry příslušnosti [15]	23
Obr. 14: Vývojový diagram pro načítání obrazových dat	25
Obr. 15: Ukázka testovaných medicínských obrazů, nahoře obrazy datasetu č.1, uprostřed obrazy datasetu č.2, dole obrazy datasetu č.3	26
Obr. 16: Vývojový diagram aplikace šumu	26
Obr. 17: Vývojový diagram implementace nehierarchických metod shlukové analýzy.....	27
Obr. 18: Vývojový diagram implementace nehierarchických metod shlukové analýzy.....	27
Obr. 19: Vývojový diagram pro evaluaci výstupních obrazů shlukové analýzy.....	28
Obr. 20: Aplikace gaussovského šumu na CT řez: 1 , nativní snímek; 2 , $\mu=0$, $\sigma=0,1$; 3 , $\mu=0$, $\sigma=0,5$	30
Obr. 21: Aplikace Salt & pepper šumu na CT řez: 1 , nativní snímek; 2 , $d=0,1$; 3 , $d=0,3$	30
Obr. 22: Aplikace speckle šumu na CT řez: 1 , nativní snímek; 2 , $\mu=0$, $\sigma=3$; 3 , $\mu=0$, $\sigma=5$	31
Obr. 23: Aplikace localvar šumu na CT řez: 1 , nativní snímek; 2 , $\mu=0$, $\sigma=0,1$; 3 , $\mu=0$, $\sigma=0,3$	31
Obr. 24: Algoritmus metody k-means	33
Obr. 25: Regionální segmentace obrazu na základě metody k-means pro 3 shluky	33
Obr. 26: Funkce příslušnosti pro jednorozměrné datové sady. Vlevo tvrdá klasifikace umožňující binární rozhodování o příslušnosti, vpravo odstupňovaná funkce příslušnosti využívaná u fuzzy c-means [23]	34
Obr. 27: Algoritmus metody fuzzy c-means	35

Obr. 28: vlevo nativní obraz, vpravo obraz po regionální segmentaci metodou fuzzy c-means pro 3 shluky	35
Obr. 29: Segmentovaný obraz metodou fuzzy c-means (vlevo), graf míry příslušnosti jednotlivých pixelů ke shlukům (vpravo).....	36
Obr. 30: Nativní snímek (vlevo), Segmentovaný obraz metodou fuzzy c-means (uprostřed), graf míry příslušnosti jednotlivých pixelů ke shlukům (vpravo)	36
Obr. 31: (a) – nativní snímek, (b) – segmentovaný snímek metodou k-means pro 3 shluky, (c) – segmentovaný snímek metodou k-means pro 5 shluků, (d) – segmentovaný snímek metodou k-means pro 8 shluků.....	37
Obr. 32: (a) – nativní snímek, (b) – nativní snímek vybrané oblasti, (c) – segmentovaný snímek metodou k-means pro 2 shluky, (d) – segmentovaný snímek metodou k-means pro 4 shluky	38
Obr. 33: (a) – nativní snímek, (b) – nativní snímek vybrané oblasti, (c) – segmentovaný snímek metodou fuzzy c-means pro 2 shluky, (d) – segmentovaný snímek metodou fuzzy c-means pro 3 shluky, (e) – segmentovaný snímek metodou fuzzy c-means pro 4 shluky, (f) – segmentovaný snímek metodou fuzzy c-means pro 5 shluků.....	38
Obr. 34: (a) – implementace gaussovského šumu, ($\mu=0$, $\sigma=0,1$), (b) – segmentovaný snímek metodou k-means pro 3 shluky, (c) – segmentovaný snímek metodou k-means pro 5 shluků, (d) – implementace gaussovského šumu, ($\mu=0$, $\sigma=0,3$), (e) – segmentovaný snímek metodou k-means pro 3 shluky, (f) – segmentovaný snímek metodou k-means pro 5 shluků.....	39
Obr. 35: (a) – implementace salt&pepper šumu, ($d=0,1$), (b) – segmentovaný snímek metodou k-means pro 3 shluky, (c) – segmentovaný snímek metodou k-means pro 8 shluků, (d) – implementace salt&pepper šumu, ($d=0,3$), (e) – segmentovaný snímek metodou k-means pro 3 shluky, (f) – segmentovaný snímek metodou k-means pro 8 shluků.....	39
Obr. 36: (a) – implementace localvar šumu, ($\mu=0$, $\sigma=0,1$), (b) – segmentovaný snímek metodou fuzzy c-means pro 2 shluky, (c) – segmentovaný snímek metodou fuzzy c-means pro 6 shluků, (d) – implementace localvar šumu, ($\mu=0$, $\sigma=0,3$), (e) – segmentovaný snímek metodou fuzzy c-means pro 2 shluky, (f) – segmentovaný snímek metodou fuzzy c-means pro 6 shluků	40
Obr. 37: (a) – nativní snímek, (b) – implementace salt&pepper šumu, ($d=0,3$), (c) – segmentovaný snímek metodou fuzzy c-means pro 2 shluky, (d) – segmentovaný snímek metodou fuzzy c-means pro 3 shluky	40
Obr. 38: (a) – implementace gaussovského šumu, ($\mu=0$, $\sigma=0,3$), (b) – segmentovaný snímek metodou fuzzy c-means pro 3 shluky, (c) – implementace gaussovského šumu ($\mu=0$, $\sigma=0,4$), (d) – segmentovaný snímek metodou fuzzy c-means pro 5 shluků	41

Obr. 39: Vlevo nativní snímek, uprostřed segmentovaný snímek s nastavením 8 shluků, vpravo extrakce cévního zásobení dolních končetin	42
Obr. 40: Vlevo nativní snímek, uprostřed segmentovaný snímek s nastavením 5 shluků, vpravo extrakce cévního zásobení mozku	42
Obr. 41: Křivka EKG záznamu [24]	43
Obr. 42: Vlevo fyziologická křivka EKG, vpravo extrakce části křivky	43
Obr. 43: Vlevo segmentace křivky metodou k-means s nastavením 2 shluků, vpravo segmentace křivky metodou k-means s nastavením 3 shluků	44
Obr. 44: Segmentace křivky metodou k-means pro extrakci ST úseku	44
Obr. 45: Vlevo EKG záznam srdeční arytmie, vpravo segmentace křivky pro 4 shluky.	45
Obr. 46: Závislost počtu shluků na kvalitě segmentace obrazu.	46
Obr. 47: Srovnání metod k-means a fuzzy c-means na jednotlivých typech šumu.....	47
Obr. 48: Graf míry difference mezi metodami k-means a fuzzy c-means.....	47
Obr. 49: Grafy míry ovlivnění a zkreslení obrazu na základě zvolení typu šumu	48
Obr. 50: Srovnání míry ovlivnitelnosti gaussovským šumem dle jednotlivých datasetů	48
Obr. 51: Graficko-uživatelské prostředí.....	49
Obr. 52: Načtení obrazu do graficko-uživatelského prostředí	49
Obr. 53: Aplikace šum	50
Obr. 54: Volba metody k-means	50
Obr. 55: Volba metody fuzzy c-means.....	51

Seznam tabulek

Tab. 1: Testovaná obrazová data.....	25
Tab. 2: Exemplární nastavení 20 hodnot s denzitou (d) šumu salt & pepper a krokem 0,0125.....	29
Tab. 3: Výsledné průměrné korelace metod pro jednotlivé šumy.....	46
Tab. 4: Výsledné průměrné střední kvadratické chyby metod pro jednotlivé šumy	46

Úvod

Lékařské zobrazovací systémy jsou v posledních letech nedílnou součástí novodobé medicíny. Moderní přístroje jsou vytvořeny na základě pokročilých technologií, a to jak při samotném získávání obrazu, tak i následně při jeho zpracování. Problém nastává při nečekaném narušení obrazu parazitním šumem, který může vzniknout již v elektronice daného přístroje nebo v závislosti na jevech způsobených vnějším prostředím.

Shluková analýza slouží k rozdělení či zařazení sady neoznačených prvků, tak že prvky jejichž příznaky vykazují signifikantní podobnost náleží do stejné skupiny, na rozdíl od prvků mimo tuto skupinu, kde je výsledná podobnost minimální. V našem případě jsou těmito prvky pixely v obraze.

Práce se zaměřuje na metody shlukové analýzy aplikované na sady obrazových medicínských dat. Cílem je především zjištění a vyhodnocení výsledků použitých nehierarchických metod k-means a fuzzy c-means v závislosti na aditivním šumu. Primárním rozdílem mezi metodami je přiřazování prvků k jednotlivým shlukům s rozdílnou mírou příslušnosti.

V praktické části práce je tvorba aplikace v prostředí MATLAB, která zahrnuje design, implementaci a dynamické testování metod shlukové analýzy v závislosti na aditivním šumu. Následuje popis jednotlivých generátorů umělého šumu a nastavení jejich vstupních parametrů s konkrétními příklady. Nedílnou součástí práce je samotný popis daných metod, jejich vnitřních struktur a nastavení počátečních shluků. Testování na reálných medicínských datech je vyobrazeno a popsáno v 7. kapitole. Na tomto základě bylo navrženo testování za účelem modelování a extrakce tkání, které je popsáno v 8. kapitole.

Ačkoliv je práce z největší části koncipovaná na zpracování obrazu, zabývá se i implementací metod shlukové analýzy pro analýzu 1D medicínských signálů s cílem možnosti dekompozice jednotlivých EKG segmentů.

Závěrečná část práce je věnována evaluaci testovaných dat. Výsledné obrazy byly za pomoci korelačního koeficientu a střední kvadratické chyby vyhodnoceny a vyneseny do jednotlivých grafů. Analyzováno bylo především srovnání jednotlivých metod, ovlivnění obrazu dle aplikace jednotlivých šumů a vliv počtu shluků na kvalitu segmentace obrazu. Závěrem byla pro usnadnění a názornou ukázkou celková práce přenesena do graficko-uživatelského prostředí.

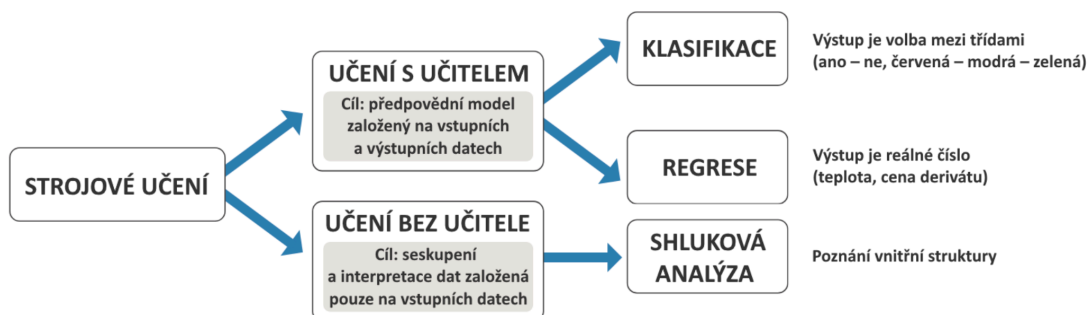
1 Strojové učení

Schopnost učení je jedním z nejdůležitějších aspektů inteligence. Převedení této síly na stroje zní jako obrovský krok směrem k jejich větší inteligenci. Strojové učení je ve skutečnosti oblast, která dělá největší pokrok v umělé inteligenci. V dnešní době hraje klíčovou roli v oblasti statistiky, těžby dat a umělé inteligence. Zabývá se algoritmy a technikami, které umožňují danému počítačovému systému přizpůsobit se okolnímu prostředí za pomoci učení. Tím rozumíme, že spuštěné algoritmy jsou schopné umožnit počítači naučit se provádět úlohy z dané sady dat namísto jejich výslovného naprogramování. [1, 2]

„Počítačový program říká, že se naučí provádět úkol T od zkušenosti E , jestliže jeho výkon na úkolu T , měřený ukazatelem výkonu P , se v průběhu času zlepšuje se zkušenostmi E .“ [2]

Představme si umělého hráče hrajícího šachy úkol T , který na základě sledování partií nebo přímou hrou se soupeřem zkušenost E zlepšuje své dosavadní výsledky měrné procentem výkonu P s rostoucími zkušenostmi E .

V tomto okamžiku mluvíme o dvou obecných kategoriích algoritmů strojového učení, a to o algoritmech pro učení s učitelem nebo bez učitele. Hlavní rozdíl mezi oběma přístupy spočívá v tom, jak do našeho algoritmu přinášíme tréninkové příklady, v jakém algoritmu je používáme a na druhu problému, který řeší. [1, 2]



Obr. 1: Dělení strojového učení [2]

1.1 Učení s učitelem

V případě učení s učitelem lze algoritmus strojového učení vnímat jako proces, který musí přeměnit určitý vstup na požadovaný výstup. Proces učení se pak musí naučit, jak přeměnit každý možný vstup na správný požadovaný výstup. V příkladu o umělém šachovém hráči by náš vstup byl určitý stav šachovnice a výstup by byl v této situaci nejlepší možný pohyb. V závislosti na typu výstupu dělíme dále učení s učitelem na podkategorie klasifikace a regrese. [1, 2]

1.1.1 Klasifikace

Pokud výstupní hodnota náleží konečné množině, můžeme hovořit o klasifikaci. Cílem je třídít vlastnostmi popsané objekty do podmnožin, které mohou či nemusí mít předem danou definici. Klasifikace je zpravidla jedním z nástrojů předcházejícím zpracování dat, které se potýká s chybějícími či odlehlými hodnotami a jejich následnými úpravami. [1, 2]

1.1.2 Regrese

Je častým nástrojem pro predikci a prognózu. Pokud je výstupní hodnota kontinuálním číslem, například pravděpodobností, hovoříme o regresi. Popisuje vztah dvou proměnných, a to závislých a nezávislých, vytvořením funkce je přímka, která predikuje hodnoty závislé proměnné jako funkci jedné nebo více proměnných nezávislých. [1, 2]

1.2 Učení bez učitele

Existuje druhá kategorie algoritmů strojového učení nazvaná učením bez učitele. V tomto případě musí být výukové příklady pouze vstupem do algoritmu, nikoliv však požadovaným výstupem. Typickým příkladem jsou algoritmy shlukování, kde se naučíme najít podobné instance nebo skupiny instancí (shluků). Téma bude podrobně rozebráno ve třetí kapitole. [1, 2]

2 Podobnost objektů

Ke stanovení podobnosti objektů jsou nejčastěji používány míry podobnosti, nepodobnosti či metriky vzdálenosti. Vzdálenost dvou bodů lze zaznamenat do symetrické čtvercové matice s nulami na hlavní diagonále.

Při shlukování poté mohou být pro zjištění distance dvou shluků použita kritéria minimální vzdálenosti dvojice objektů, jejich maximální vzdálenosti a průměrné vzdálenosti. Rozhodnutí o konečném počtu shluků vychází jak z teoretických vzdáleností shluků, tak i z rozhodnutí, zda chceme shlukovat jednotlivé objekty, proměnné nebo kategorie dat. [6]

2.1 Metriky vzdálenosti mezi dvěma obrazy s kvantitativními příznaky

Vztahy mezi objekty lze vyjádřit pomocí metrik vzdálenosti. Jejich společnou vlastností je, že maximální hodnotu dosahují dva objekty, které jsou nejvíce odlišné a objekty identické mají vzdálenost nulovou. Použití konkrétní metriky vzdálenosti závisí na řešené úloze.

Také lze při výběru možné metriky použít i další dílčí kritéria, jako je charakter rozložení dat a výpočetní nároky. Obecně nelze doporučit vhodný postup pro výběr určité metriky na daný typ úlohy. Vzdálenost budeme dále označovat písmenem D . [3, 4]

Metriky vzdálenosti musí dodržovat následující podmínky:

- Pokud jsou objekty stejné, vzdálenost mezi nimi se rovná 0.
- Pokud se $x = y$, potom $D(x, y) = 0$ platí, že mezi rozdílnými objekty je vzdálenost kladná.
- Pokud $x \neq y$, potom $D(x, y) > 0$ platí, že vzdálenost objektu x od y je totožná, jako vzdálenost y od x .

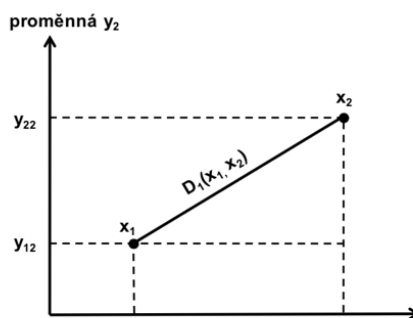
Pokud máme součet $D(x, y) + D(y, x)$, musí být vždy $\geq D(x, z)$, platí zde trojúhelníková nerovnost. [3]

2.1.1 Euklidovská vzdálenost

Jde o jednu z nejpoužívanějších měr vzdálenosti. Je založena na Pythagorově větě a znázorňuje Euklidovskou vzdálenost dvou objektů v prostoru o dvou proměnných.

Je to metrika zřejmě s nejnázornější geometrickou interpretací, geometrickým místem bodů s toutéž Euklidovou vzdáleností od daného bodu je koule. [3, 4]

$$D_2(x_1, x_2) = \sqrt{\sum_{j=1}^P |y_{1j} - y_{2j}|} \quad (1)$$



Obr. 2: Znázornění Euklidovské vzdálenosti v prostoru [3]

2.1.2 Manhattan vzdálenost

Také nazývaná v angličtině „city-block metrika“, protože svým výpočtem ve dvourozměrném prostoru připomíná vzdálenost, kterou urazí automobil jedoucí z jednoho místa do druhého v pravoúhlém prostředí. Je definována tímto vztahem. [3, 4]

$$D_2(x_1, x_2) = \sqrt{\sum_{j=1}^P |y_{1j} - y_{2j}|} \quad (2)$$

2.1.3 Minkowského vzdálenost

Podle zvoleného koeficientu může odpovídat euklidovské nebo manhattanské metrice. Se stoupajícím koeficientem umocňování stoupá významnost větších rozdílů. To znamená, že čím větší mocnina, tím větší důraz na velké rozdíly mezi příznaky. [3, 4]

$$D_3(x_1, x_2) = \left[\sum_{j=1}^P |y_{1j} - y_{2j}|^\lambda \right]^{1/\lambda} \quad (3)$$

2.2 Metriky podobnosti dvou obrazů s kvantitativními příznaky

2.2.1 Skalární součin

Skalární součin definujeme mezi dvěma sloupcovými vektory. Jeho výsledkem je reálné číslo. V Euklidovském prostoru je pro dvojici vektorů definován tímto vztahem. [4]

$$D_3(x_1, x_2) = x_1^T \cdot x_2 = \sum_{i=1}^n x_{1i} \cdot x_{2i} \quad (4)$$

2.2.2 Tanimotova metrika podobnosti

Tanimotova podobnost je použitelná pouze pro binární proměnné. Pohybuje se v rozmezí od 0 do 1 (kde 1 je nejvyšší podobnost). [4]

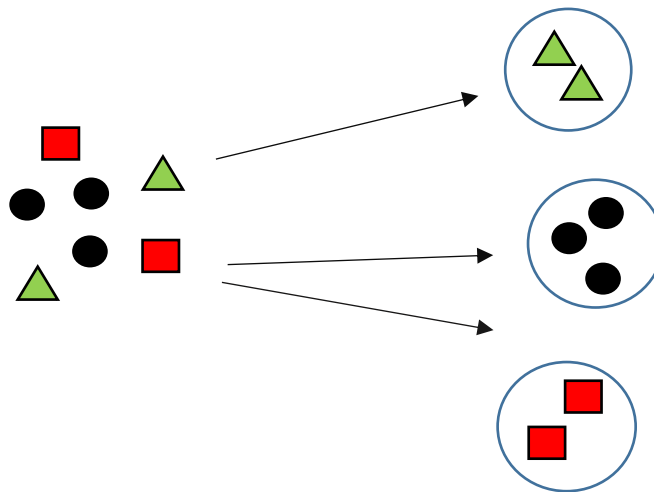
$$D_T(x_1, x_2) = \frac{x_1^T x_2}{\|x_1\|^2 + \|x_2\|^2 - x_1^T x_2} \quad (5)$$

3 Shluková analýza

Shluková analýza je metodou seskupování sady různorodých neoznačených prvků takovým způsobem, že prvky ve stejné skupině jsou si navzájem maximálně podobné, na rozdíl od prvků mimo tuto skupinu, kde je tato podobnost minimální. Přesnost většiny těchto algoritmů se odvíjí nejen podle jejich tvarů a četnosti, ale především prostorovými vztahy a vzdáleností jednotlivých shluků mezi sebou.

Jednotlivé shluky lze tedy chápat jako sbírku objektů, které jsou si mezi sebou podobné a zároveň jsou odlišné od objektů patřících k jiným skupinám. Shluky mohou být rozdělovány, spojovány či vzájemně překrývány.

Shlukovou analýzu řadíme do kategorie „učení bez učitele“, uplatnění nalézáme často v odvětvích jako je statistika, strojové učení, bioinformatika nebo právě analýza a zpracování obrazu. [8, 12]



Obr. 3: Rozdělení objektů do shluků [vlastní]

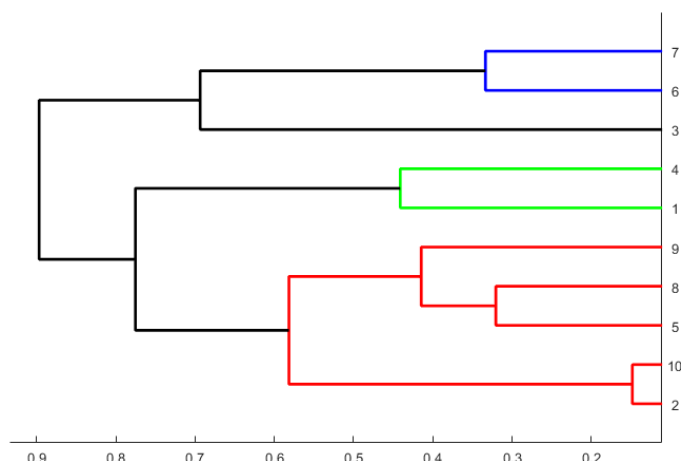
3.1 Hierarchické metody

Hierarchické metody vytváří posloupnost oddílů s jediným hlavním shlukem obsahujícím všechny dané objekty umístěným na jedné straně a samostatnými jednoprvkovými shluky umístěné na straně druhé. Každou ze středních úrovní lze považovat za kombinaci dvou shluků z úrovně nižší (nebo rozdělení shluků z další vyšší úrovně).

Výsledky algoritmu hierarchického shlukování lze graficky zobrazit jako graf, který se nazývá dendrogram neboli také binární strom. Tento graf nám zobrazuje spojitost a průběh jednotlivých mezilehlých shluků. Každý z uzlů grafu představuje shluk. Svislý směr v grafu představuje vzdálenost mezi shluky.

Nevýhodou této metody je, že při učinění rozdělní či spojení jednotlivých shluků již nemůžeme zpětně docílit jejich navrácení do původního stavu a tím případně daný algoritmus pozměnit nebo vylepšit. Existují dva základní přístupy k vytváření hierarchického shlukování: [6, 9]

- Aglomerativní (počínaje jednotlivými prvky a jejich agregací do shluků),
- Divizní (počínaje hlavní množinou objektů, která je následně dělena na podmnožiny).



Obr. 4: Dendrogram [vlastní]

3.1.1 Aglomerativní metody

Metoda vychází z n jednoprvkových objektů, které jsou každý jednotlivým shlukem. Postupně spojujeme objekty po dvojicích, a to v pořadí od shluků s maximální podobností k podobnosti minimální. Výsledkem aglomerativní metody je právě jeden shluk. [11, 12]

Při použití aglomerativní metody postupujeme po krocích, v každém jednotlivém kroku se spojují dva nejpodobnější shluky. Tuto podobnost vyjadřujeme za pomoci některých z níže zmíněných kritérií, dle kterých se vybírají shluky s největší podobností či nepodobností. [12]

- **Metoda nejbližšího souseda**

Vzdálenost je vyjádřena minimální vzdáleností libovolného bodu ve shluku vůči bodu v jiném shluku. Takto nalezené body se spojují nejkratší spojkou. Tato metoda není vhodná pro shluky o stejné vzdálenosti od již vytvořených shluků, kde může docházet ke kolizím. [17]



Obr. 5: Metoda nejbližšího souseda [13]

- **Metoda nejvzdálenějšího souseda**

Metoda seskupuje naopak maximálně vzdálené skupiny bodů na základě největší vzdálenosti mezi body ve dvou skupinách. Je citlivá na odlehle hodnoty, které mohou pozměnit výsledné shlukování. [17]



Obr. 6: Metoda nejvzdálenějšího souseda [13]

- **Metoda průměrné vazby**

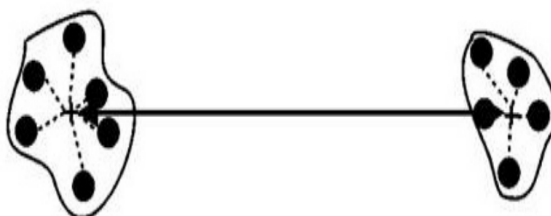
Podobnost dvou shluků se počítá jako průměr vzdáleností mezi každými dvěma objekty patřícími do dvou různých shluků. Nejpodobnější jsou si shluky s nejmenší průměrnou vzdáleností. [13, 17]



Obr. 7: Metoda průměrné vazby [13]

- **Wardova metoda**

Hlavním účelem není zjištění vzdálenosti shluků, ale minimalizování ztráty důležitých dat při vytváření nových shluků. Metoda je založena na ztrátě informací, která vzniká při shlukování. Kritériem pro shlukování je celkový součet druhých mocnin odchylek každého objektu od těžiště shluku, do kterého náleží. [17]



Obr. 8: Wardova metoda [13]

- **Centroidová metoda**

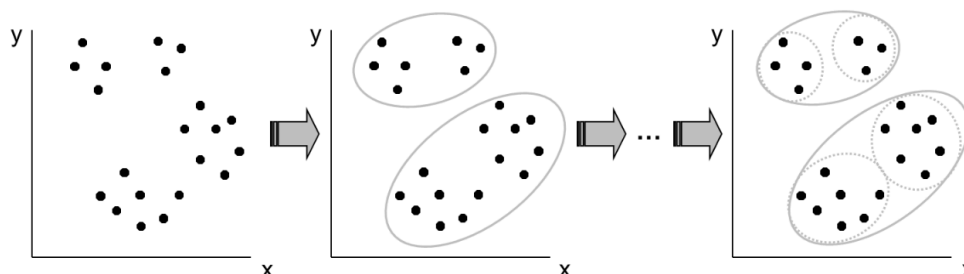
Metoda je soustředěna na výpočet vzdáleností těžišť shluků, ty jsou pak následně slučovány dle vzdálenosti mezi těžišti od nejmenších. Pro výpočet je využívána Euklidovská metrika vzdálenosti. [14]



Obr. 9: Centroidová metoda [14]

3.1.2 Divizní metody

Divizní metody neboli dělení od největšího shluku, předpokládá, že všechny objekty na začátku tvoří jediný shluk, který následně rozdělí na dva shluky s minimální podobností. Tento postup se rekurzivně opakuje, dokud nedojdeme k existenci samostatného shluku pro každé z pozorování. [10, 11, 12]



Obr. 10: Znáznornění postupu tvoření shluků za pomoci divizních metod [14]

3.2 Nehierarchické metody

Často nazývané také metodami rozkladu, které se uplatňují na místech, kde data netvoří hierarchickou strukturu. Algoritmu je přiřazena konkrétní oblast dat podle předem daných podmínek, prvotní rozklad se již dále nedělí.

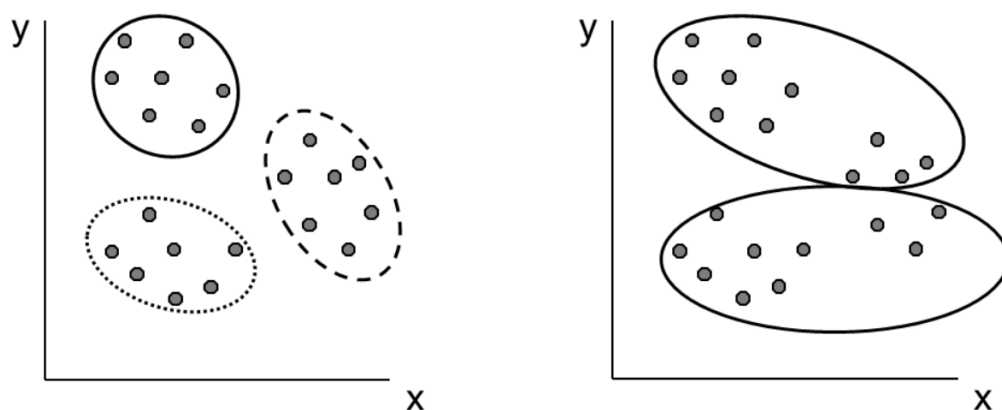
Přidělení k jednotlivým shlukům je buď dané, nebo se optimalizuje za pomoci míry příslušnosti jednotlivých objektů ke shlukům. Nehierarchické metody mají výhodu v aplikacích, které obsahují velké množství dat a použití hierarchických metod by zde bylo výpočetně velmi náročné. [11, 12]

3.2.1 Metoda k-means

Tato metoda je především vhodná pro vytvoření menšího počtu jednotlivých shluků z velké sady dat. Je náchylná na odlehlé hodnoty, které je nutné předem normalizovat. Metoda k-means byla pojmenována podle počtu shluků k , do kterých budou data seskupeny, počet musí být před analýzou specifikován společně s počátečním odhadem o členství objektů všech shluků.

Potenciální nevýhodou metody je, že pokaždé dochází k jinému výslednému řešení, neboť neexistuje univerzální metoda, která by identifikovala počáteční objekty a počty shluků, ty jsou proto zadány předem uživatelem nebo zvoleny náhodně do daných skupin, a i nadále v nich zůstávají.

Metoda neumožňuje žádnými prostředky pro již přiřazené objekty následně zpětnou relokaci, proto jsou daná rozdělení dat spouštěna většinou náhodně a opakovaně za účelem nalezení optimálního výsledku. Způsoby dělení obecně vytváří k oddíly datových sad s n objekty, každá oblast představuje shluk, kde $k \leq n$. Tato metoda bude podrobněji rozebrána v praktické části práce. [11, 12]



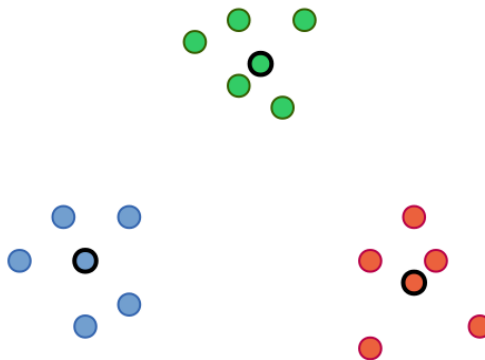
Obr. 11: „Ukázka rozdělení objektů do shluků nehierarchickou metodou *k-means*. Výsledek je ovlivněn volbou počtu shluků. Vlevo: počet shluků $k=3$ je dobrá volba; vpravo: počet shluků $k=2$ je špatná volba.“ [14]

3.2.2 Metoda PAM

Nazývaná také metoda *k-medoidů*. Vychází z počátečního rozdělení objektů do k shluků, stejně jako předchozí metoda *k-means*, avšak oproti ní, této metodě nevadí odlehlé hodnoty a šum v datech.

Po vytvoření shluků je každému z nich přiřazen medoid, který představuje jejich konkrétní objekt. Prvotní medoid je nastaven tak, aby součet vzdáleností objektů od něj byl minimální.

Následně přezkoumáváme objekty, které jsou nejbližší danému vlastnímu medoidu. Takto jsou zachovány v původním shluku. Vzdálenější objekty přemísťujeme do shluku, k jehož medoidu mají nejbližší vzdálenost. V další fázi se zjišťuje nejefektivnější rozložení k bodů, přidávají se zatím nezařazené body. [11, 12]



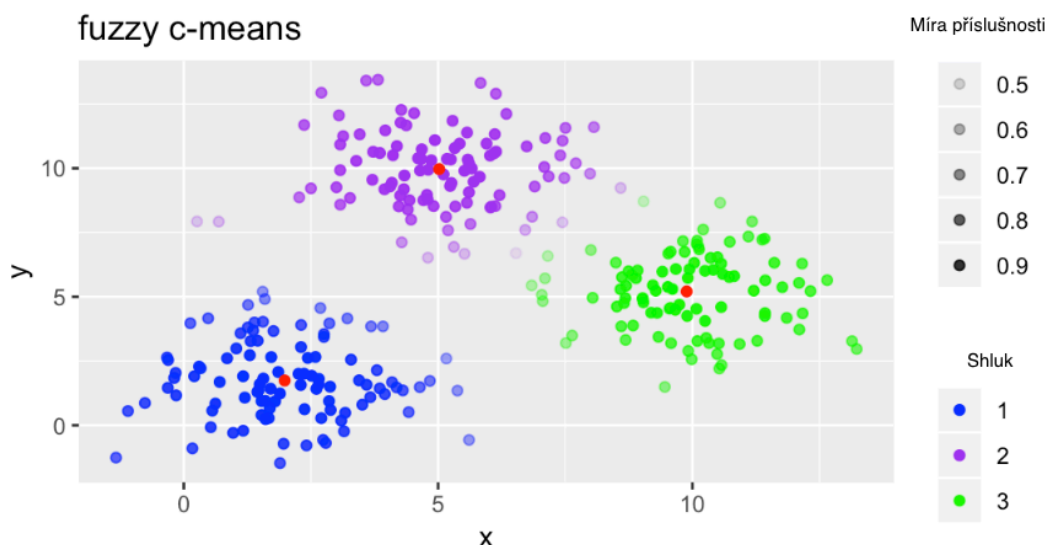
Obr. 12: Rozdělení shluků pro metodu PAM [vlastní]

3.2.3 Fuzzy c-means

Na rozdíl od předešlých shlukovacích metod s takzvaným tvrdým nastavením přístupu, což znamená, že daný objekt buď do shluku patří, nebo nepatří. Fuzzy c-means umožňuje, aby objekty mohly patřit do více shluků zároveň.

Každému objektu je přiřazena váha příslušnosti k určitému shluku, nikoliv přiřazení každého objektu k jeho jedinečnému shluku. Objekty vně shluku se tudíž mohou vyskytovat s menší četností než objekty nacházející se ve středu shluku.

Aplikace metody je využívána v dolování dat, rozpoznávání vzorů, klasifikaci a segmentaci obrazu. Podrobný popis metody bude rozebrán v praktické části práce. [11, 12]



Obr. 13: Model aplikování metody fuzzy c-means, volba počtu shluků, nastavení míry příslušnosti [15]

3.2.4 Metoda CLARA

Algoritmus CLARA (Clustering Large Applications) je rozšířením metody PAM u velkých datových sad, kde objekty dosahují řádu tisíců. Zamýšlí snížit výpočtový čas v případě velkého souboru dat. Jako téměř každý algoritmus nehierarchických metod vyžaduje uživateli určit příslušný počet shluků, které mají být vytvořeny. To lze odhadnout pomocí funkce předem nadefinovaných funkcí.

Namísto zjištění medoidů pro celou sadu dat, metoda CLARA vezme v úvahu malý vzorek dat s pevnou velikostí a aplikuje algoritmus PAM pro vytvoření optimální sady medoidů pro vzorek. Kvalita výsledných medoidů se měří podle průměrných rozdílů mezi každým objektem v celé sadě dat a medoidem jeho shluku.

CLARA opakuje procesy vzorkování a shlukování předem stanoveným počtem opakování, aby se minimalizovalo zkreslení vzorků. Konečné výsledky shlukování odpovídají souboru medoidů. Pokud je kvantita zpracovávaných dat větší než shlukovaný vzorek, pak metoda nedosahuje přesného výsledku. [16]

3.2.5 Metoda CLARANS

Metoda CLARANS (Clustering Large Application based on randomised Search) vychází z obdoby metody CLARA. Oproti PAM a CLARA její přednosti spočívají v časové složitosti. Počáteční fází je určení, do jaké vzdálenosti od medoidu se budou přilehlé objekty prohledávat, a to s každou nadcházející iterací. Zároveň se určuje i počet těchto iterací. Samotná metoda se dá jednoduše popsat jako prohledávání dendogramu, kde každý uzel může být ten pravý medoid, kterým je zajištěno optimální rozložení shluků.

Metoda začíná náhodně zvoleným uzlem a pro předem zvolenou vzdálenost prohledávání prozkoumá sousední uzly. Pokud je nalezen lépe vyhovující uzel, tak ho vybere a pokračuje z něj. Pokud takový uzel nenajde, označí stávající uzel za lokální minimum. Konečným stavem je, pokud se dosáhne předem daného počtu iterací. [17, 18]

3.2.6 Bagged shlukování

Jedná se o poměrně novou metodu, která je schopná kombinovat hierarchický a nehierarchický přístup. Centrální myšlenkou je stabilizace nehierarchických metod, především k-means, neboť jeho výsledky jsme schopni pouze lokálně optimalizovat.

Bagged shlukování může být aplikováno dvěma různými způsoby. Prvním z nich je předběžná fáze snižování složitosti pro hierarchické metody a druhým kombinace procedur pro nehierarchické shlukování. [17, 18]

4 Experimentální část

Na základě znalostí získaných z teoretické části práce je v praktické části přistoupeno k realizaci cíle bakalářské práce, kterým je design a implementace algoritmů nehierarchických metod shlukové analýzy na reálná obrazová medicínská data a jejich následné porovnání. Zpracování proběhlo v programovacím jazyce MATLAB. Aplikace byla testována na třech skupinách datasetů, ve kterých bylo vždy po deseti obrazech, viz Tab. 1. Realizace aplikace je rozvržena do následujících dílčích částí.

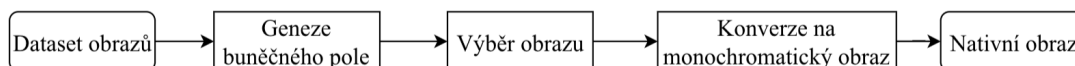
4.1 Načtení medicínských obrazů

Skript pro načtení reálných medicínských obrazů do prostředí MATLAB funguje na principu výběru daného datasetu obrazů, který je skenován a obrazy z něj jsou načítány do buněčného pole. Následně je vybrán jeden obraz, který bude použit pro testování, viz příloha I.

Tento obraz se převede z barevného RGB modelu na monochromatický, dojde tedy k odfiltrování barevné složky obrazu při zachování jeho jasových vlastností, obraz bude následně zobrazován v odstínech šedi. Pro tuto transformaci byla použita funkce, která je součástí prostředí MATLAB.

$$Y = [(0,3 \cdot R) + (0,59 \cdot G) + (0,11 \cdot B)] \quad (6)$$

Princip funkce spočívá ve vytvoření váženého součtu jednotlivých složek RGB modelu. Pro každou složku je dán koeficient, kterým se složka násobí, viz rovnice (6). Jednotlivé koeficienty nám procentuálně vyjadřují příspěvek barev RGB do modelu stupně šedi. Červená barva (R) s 30 %, zelená (G) s 59 %, což je nejvyšší podíl ze všech 3 barev a modrá (B) s nejmenší hodnotou 11 %. Díky tomuto rozložení bude mít výsledný obraz správné jasové vlastnosti. V případě procentuálního převýšení červené nebo modré barvy nad zelenou bude obraz jasově utlumen. Je to dáno tím, že lidské oko je nejcitlivější právě na zelenou barvu.



Obr. 14: Vývojový diagram pro načítání a konverzi obrazových dat

Tab. 1: Testovaná obrazová data

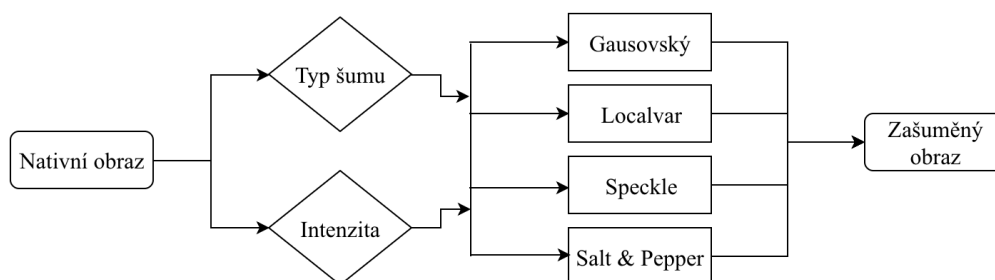
Dataset	1	2	3
Zdroj obrazů	Angiografie cévní kalcifikace	CT řezy jater	MRI kloubní chrupavky
Počet	10	10	10
Rozměr [px]	1024 x 1024	630 x 630 442 x 442	512 x 512
Barevný model	RGB	RGB	RGB



Obr. 15: Ukázka testovaných medicínských obrazů, nahoře obrázky datasetu č.1, uprostřed obrázky datasetu č.2, dole obrázky datasetu č.3

4.2 Implementace šumu

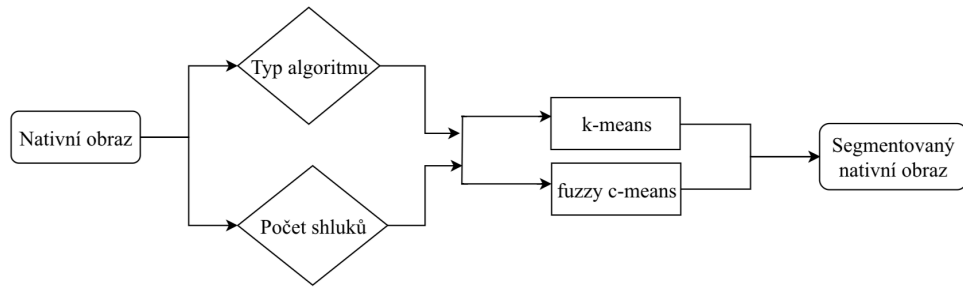
Implementace aditivního šumu u jednotlivých obrazů je realizována za pomoci předdefinované funkce MATLABu. Pro aplikaci je vybráno ze čtyř šumů, viz Obr. 16. U každého jsou nastaveny stejné vstupní parametry pro stanovení intenzity. Vstupem je nativní obraz, který po implementaci vychází jako obraz zašuměný. Tomuto tématu je věnována celá kapitola 5, kde je implementace šumu popsána detailně.



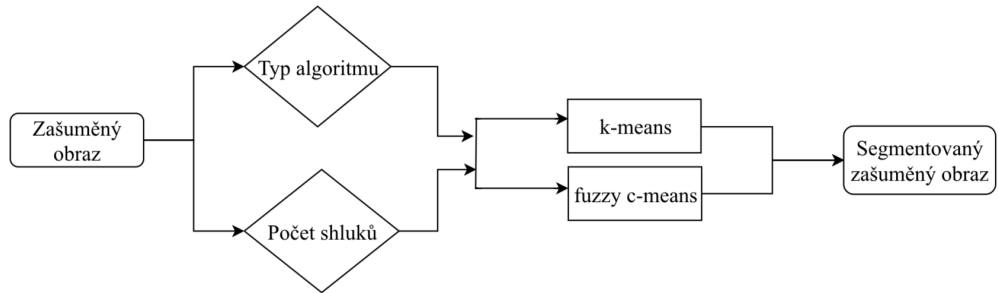
Obr. 16: Vývojový diagram aplikace šumu

4.3 Implementace algoritmu

Vstup implementace je z důvodu dalšího porovnání rozdělen do dvou dílčích částí, a to segmentace nativního a zašuměného obrazu. K dispozici jsou dvě metody nehierarchické shlukové analýzy. Segmentovaný nativní obraz je brán jako zlatý standard a je určen pro další analýzu a porovnání spolu se segmentovanými zašuměnými obrazy.



Obr. 17: Vývojový diagram implementace nehierarchických metod shlukové analýzy



Obr. 18: Vývojový diagram implementace nehierarchických metod shlukové analýzy

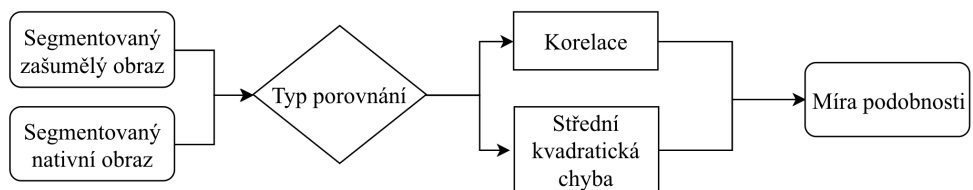
4.4 Evaluační metriky

Pro porovnání výstupu byly vybrány dvě metody. MATLAB disponuje předdefinovanou funkcí, která reprezentuje střední kvadratickou chybu. Ta svým výpočtem a zpracováním patří k jednodušším metodám pro objektivní hodnocení kvality mezi obrazy. Mse je počítána pro všechny pixely zvlášť, tudíž nemusíme brát v potaz jejich vzájemnou závislost.

$$Mse = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (7)$$

Další z metod porovnání výstupních segmentací je korelační koeficient, který je v MATLAB realizován za pomoci přednastavené funkce. Tato funkce vrací hodnotu r , která reprezentuje podobnost mezi dvěma vstupními obrazy.

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A}) (B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2) (\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \quad (8)$$



Obr. 19: Vývojový diagram pro evaluaci výstupních obrazů shlukové analýzy

5 Implementace šumu

Každý obraz obsahuje určitý druh šumu různého původu. Šum představuje parazitní část pořízeného obrazu, která zhoršuje jeho kvalitu. Šum může vznikat již v elektronice daného zařízení pořizující snímek či vnějšími vlivy jako jsou teplota, vlhkost nebo prašnost prostředí. Zhoršení a zašumění obrazu vzniká také při jeho následné úpravě nebo kompresi.

Odstranění šumu z obrazových dat je důležitou částí předzpracování pro další analýzu obrazu. Důležitým krokem při odstraňování šumu je zachování jeho původních vlastností, nemělo by tedy dojít ke zkreslení hran obrazu či jeho rozostření. [20]

Všechny syntetické šumy byly implementovány v prostředí MATLAB za pomoci předdefinované funkce šumového generátoru. U každého z nich byla nastavena variabilní intenzita šumu s cílem sledování dynamiky šumu v kontextu degradace medicínských obrazů.

Tab. 2: Exemplární nastavení 20 hodnot s denzitou (d) šumu salt & pepper a krokem 0,0125

Krok	1	2	3	4	5	6	7	8	9	10
Intenzita šumu (d)	0,0125	0,025	0,0375	0,05	0,0625	0,075	0,0875	0,1	0,1125	0,125

Krok	11	12	13	14	15	16	17	18	19	20
Intenzita šumu (d)	0,1375	0,15	0,1625	0,175	0,1875	0,2	0,2125	0,225	0,2375	0,25

5.1 Gaussovský šum

Představuje statistický šum, který má pravděpodobnost hustoty rozložení rovnající se normální distribuci Gaussova rozdělení. Ovlivňuje všechny pixely daného obrazu. Jeho amplitudové spektrum je konstantní. Je často označován jako elektronický šum, neboť může vznikat v elektronice daného zařízení, na vzniku se také může podílet teplota prostředí. U šumu je vstupním parametrem střední hodnota a rozptyl. Je popsán vzorcem:

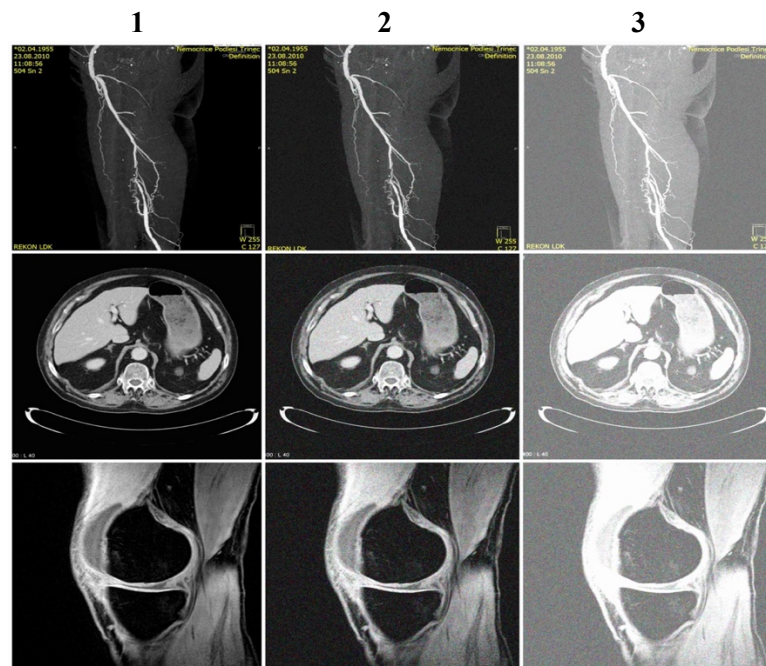
$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (9)$$

kde:

x je jasová intenzita šumu

σ je rozptyl

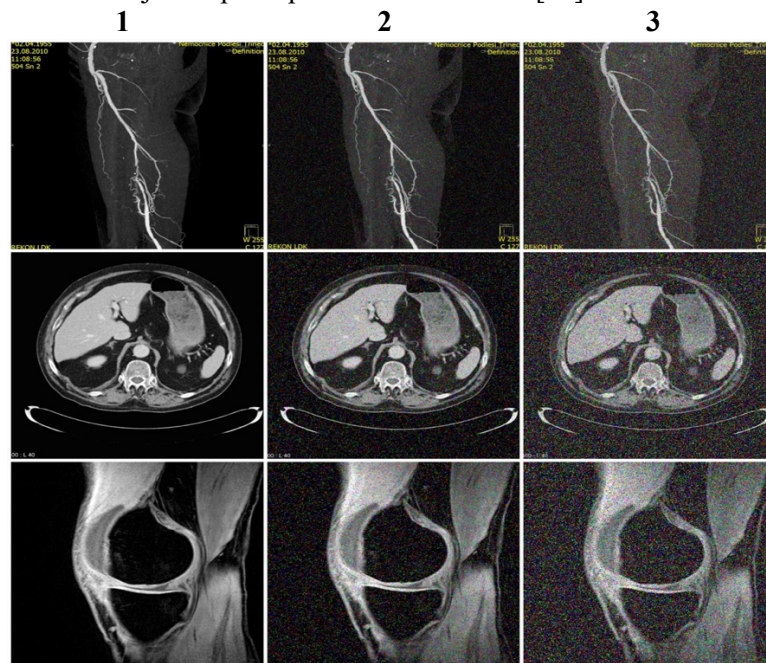
μ je střední hodnota.



Obr. 20: Aplikace gaussovského šumu na CT řez: 1, nativní snímek; 2, $\mu=0$, $\sigma=0,1$; 3, $\mu=0$, $\sigma=0,5$

5.2 Šum sůl a pepř

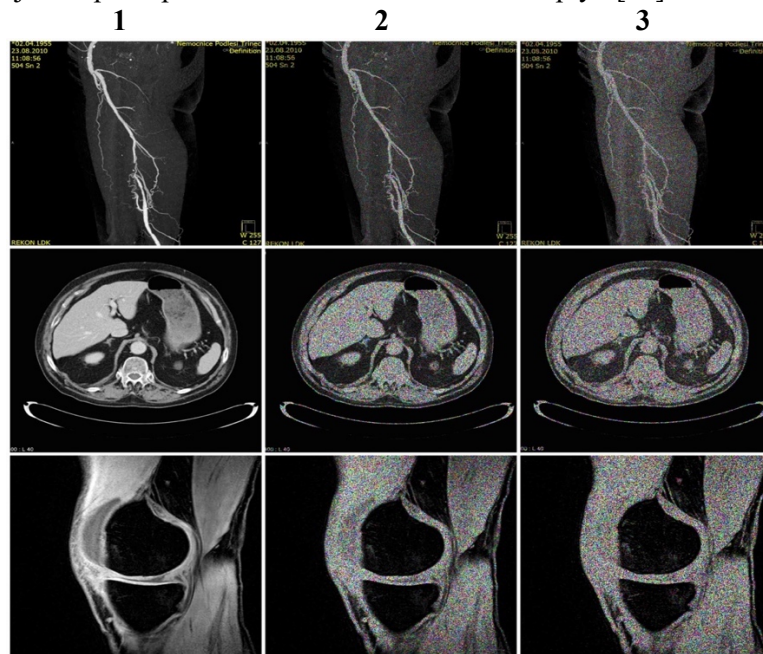
Známý také jako impulsní šum, bývá způsoben ostrým a náhlým narušením obrazového signálu. Vyskytuje se jako řídce rozložené bílé a černé pixely. Jinými slovy, část šumu typu pepř je reprezentován hodnotami dle RGB modelu blízcími se k 0 a část typu šumu sůl hodnotami co nejvyššími, až k 255. U šumu je vstupním parametrem denzita. [21]



Obr. 21: Aplikace Salt & pepper šumu na CT řez: 1, nativní snímek; 2, $d=0,1$; 3, $d=0,3$

5.3 Šum speckle

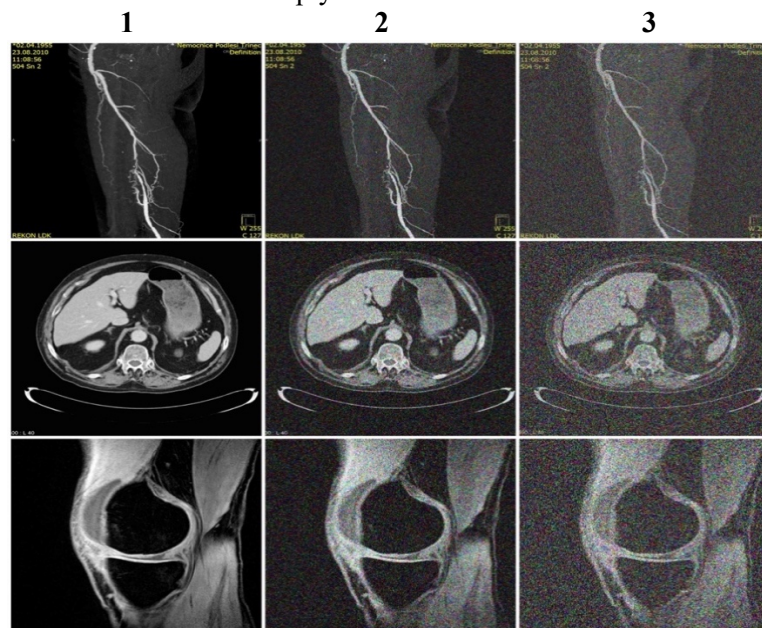
Tento typ šumu se vyskytuje nejčastěji u koherentních zobrazovacích systémů, jako jsou ultrazvukové systémy, lasery nebo magnetické rezonance. Vzniká při interferenci vracejícího se signálu odraženého od snímaného objektu zpět ke snímači. Výsledkem jsou tmavé pixely s vyšší jasovou intenzitou. U šumu je vstupním parametrem střední hodnota a rozptyl. [22]



Obr. 22: Aplikace speckle šumu na CT řez: 1, nativní snímek; 2, $\mu=0$, $\sigma=3$; 3, $\mu=0$, $\sigma=5$

5.4 Šum localvar

Jedná se o obdobný typ gaussovského bílého šumu s lokálním rozptylem hodnoty v obraze. Hodnota intenzity šumu je dána polem a musí být sejné velikosti jako vstupní matice obrazu. U šumu je vstupním parametrem střední hodnota a rozptyl.



Obr. 23: Aplikace localvar šumu na CT řez: 1, nativní snímek; 2, $\mu=0$, $\sigma=0,1$; 3, $\mu=0$, $\sigma=0,3$

6 Implementace metod shlukové analýzy na dvourozměrné signály

Pro optimální funkci metody je důležité definovat počet shluků, což je nezbytný parametr pro všechny středové shlukovací algoritmy, který zadává uživatel nebo je počet zvolen náhodně. V tomto případě je to 3, 5 nebo 8 shluků. Při nekorektním nastavení algoritmu není schopen dosáhnout konvergence a výstup regionální segmentace může být nepřesný.

Po nastavení počtu shluků je každému z nich přiděleno těžiště, ke kterému se následně přiřazují okolní objekty na základě metrik vzdálenosti. Takto přiřazené objekty se s každou další iterací přerozdělí, dokud nedojde k ustálení a všechny objekty již zůstávají na stejném místě ve správném shluku. Těžiště každého shluku tvoří aritmetický průměr všech bodů, které mu náleží. Každý z objektů je vždy blíže k vlastním shluku než ke shluku cizímu. Objekty náležící určitému shluku jsou pevně či s určitou mírou náležitosti přiřazeny ke shlukům.

Každý shluk má své unikátní číslo nazývané také indexem shluku. Tato unikátní čísla jsou generována v rámci indexační matice, pro kterou se využívá barevné kódování. Následně je tedy možné klasifikovat jednotlivé shluky dle jejich přiřazených barev.

6.1 Metoda k-means

První z metod je k-means, kterou řadíme mezi metody s tvrdým přístupem, neboť při přidělování objektů k jednotlivým shlukům můžeme stanovit pouze rozhodnutí, zda daný objekt ke shluku patří či nepatří na logické úrovni 1 nebo 0.

Algoritmus metody k-means:

1. Zvolení počtu shluků.
2. Náhodné přiřazení objektů ke shlukům.
3. Výpočet těžiště pro každý shluk.
4. Výpočet vzdálenosti všech objektů od těžiště shluku.
5. Přiřazení objektů k nejbližšímu z těžišť.
6. Opakování cyklu, dokud není přidělení shluku co nejstabilnější.

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2, \quad (10)$$

kde:

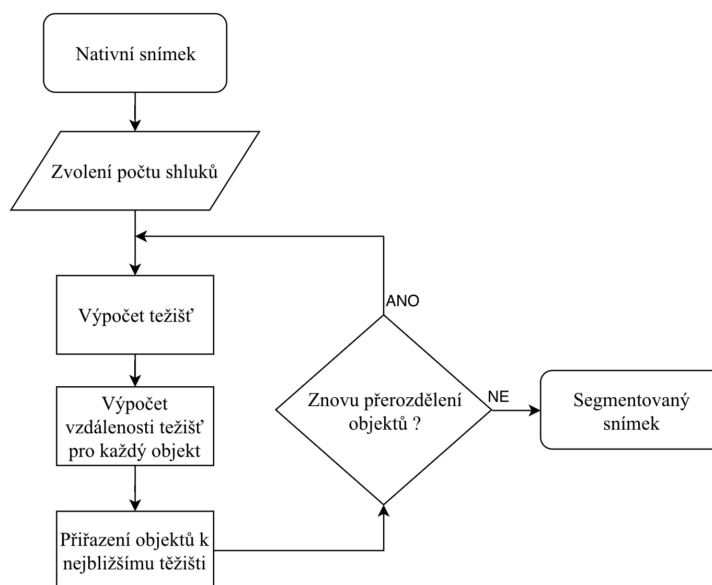
J je objektivní funkce

k je počet shluků

n je počet objektů

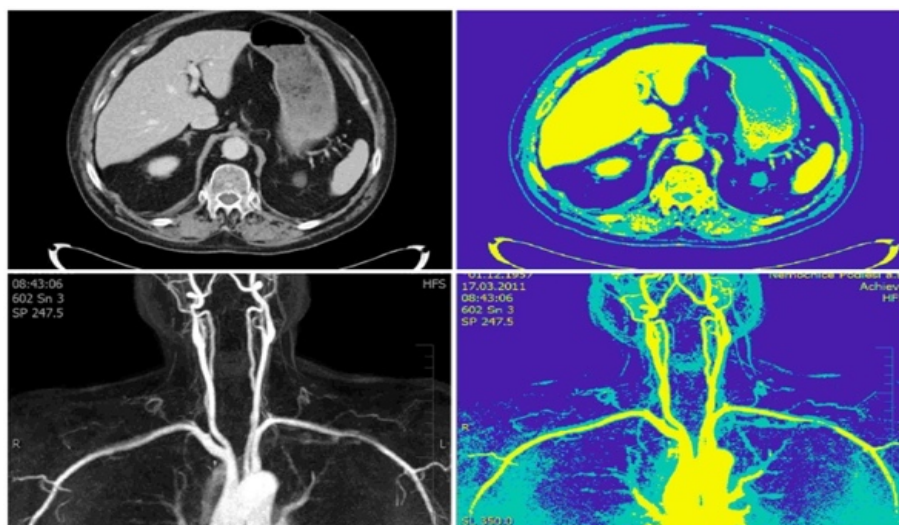
x je objekt

c je těžiště shluku.



Obr. 24: Algoritmus metody k-means

Na Obr. 25. můžeme vidět implementaci metody k-means pro regionální segmentaci s nastavením 3 počátečních shluků. Objekty v obraze s nižším stupněm jasu byly potlačeny do pozadí. Viditelné jsou především jasná místa tkání, které jsou reprezentovány shlukem žluté barvy. V případě obrazu nahoře je viditelná především oblast jater, která je hojně prokrvena, stejně jako na obrazu dole srdeční sval, aorta a hlavní tepny.

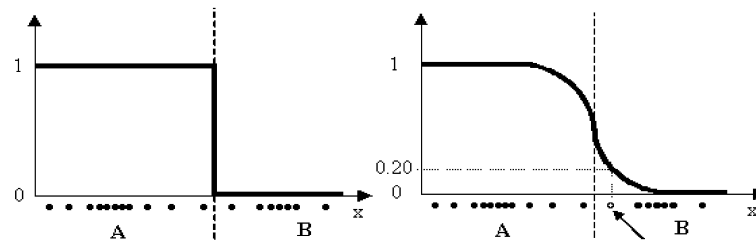


Obr. 25: Regionální segmentace obrazu na základě metody k-means pro 3 shluky

6.2 Metoda fuzzy c-means

Algoritmus metody fuzzy c-means je obdobný jako u metody k-means. Oproti předchozí metodě je ale testovaná metoda fuzzy c-means založena na fuzzy logice. Fuzzy logika je vícehodnotová logika definovaná funkcí příslušnosti prvku na intervalu $<0; 1>$, kterých může být nekonečně mnoho.

Můžeme tedy stanovit náležitost pixelů ke shlukům s určitou mírou příslušnosti. Metoda je taktéž schopna přiřadit jeden objekt k více shlukům najednou, avšak každý právě s jinou mírou příslušnosti viz Obr. 26. Vstupní počet shluků u fuzzy c-means, byl zadáván ve stejném počtu jako u metody k-means.



Obr. 26: Funkce příslušnosti pro jednorozměrné datové sady. Vlevo tvrdá klasifikace umožňující binární rozhodování o příslušnosti, vpravo odstupňovaná funkce příslušnosti využívaná u fuzzy c-means [23]

Algoritmus metody fuzzy c-means:

1. Zvolení počtu shluků.
2. Náhodné přiřazení míry příslušnosti každému z objektů.
3. Výpočet těžiště shluku.
4. Výpočet míry příslušnosti každého objektu k danému shluku.
5. Tyto dva předchozí body jsou opakovány, dokud není dosaženo optimálního řešení.

$$J_m = \sum_{j=1}^D \sum_{i=1}^N \mu_{ij}^m \|x_i^{(j)} - c_j\|^2, \quad (11)$$

kde:

D je počet objektů

N je počet shluků

m je objekt patřící současně do více než jednoho shluku zároveň

x je každý i -tý objekt

c je těžiště j -shluků

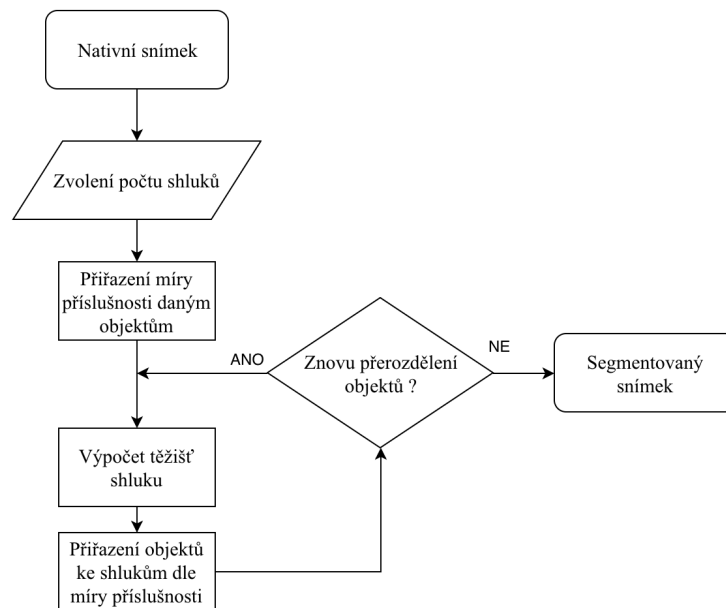
μ je míra příslušnosti objektu ke shluku.

Krok 2,3

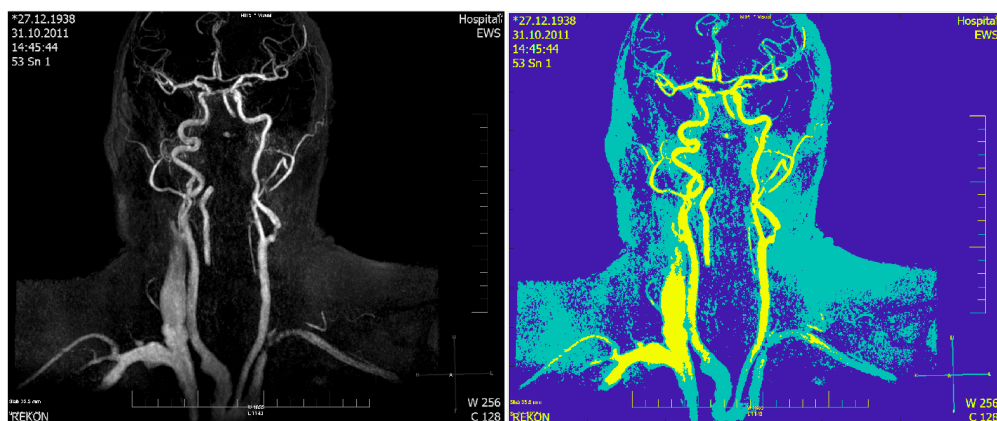
$$c_j = \frac{\sum_{i=1}^D \mu_{ij}^m x_i}{\sum_{i=1}^D \mu_{ij}^m} \quad (12)$$

Krok 4

$$\mu_{ij} = \frac{1}{\sum_{k=1}^N \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (13)$$



Obr. 27: Algoritmus metody fuzzy c-means



Obr. 28: vlevo nativní obraz, vpravo obraz po regionální segmentaci metodou fuzzy c-means pro 3 shluky

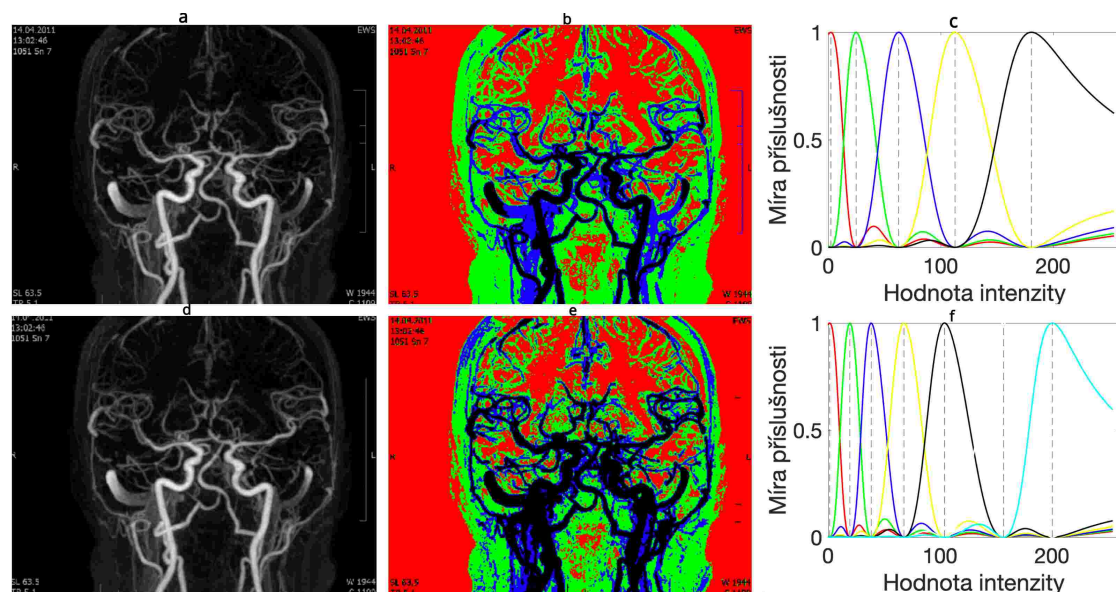


Obr. 29: Segmentovaný obraz metodou fuzzy c-means (vlevo), graf míry příslušnosti jednotlivých pixelů ke shlukům (vpravo)

Obrazová matice je rozdělena na jednotlivé pixely, každý z nich má určitou hodnotu v intervalu $<0; 255>$. V případě formátu obrazu *uint-8* rozeznáváme právě 2^8 jasových úrovní, což se rovná 256 hodnotám. Hodnota 0 je absolutně černá barva a 255 je absolutně bílá, vše mimo tyto hodnoty jsou jednotlivé odstíny šedi.

Na Obr. 29 (vpravo) je znázorněna v grafu závislost jasových intenzit pixelů na jejich příslušnosti ke shlukům. Každá z barevných křivek reprezentuje právě jeden shluk. Na ose x je hodnota intenzity jasu každého z pixelů. Na ose y jsou hodnoty reprezentující míru náležitosti pixelů určité intenzity ke shlukům. Pixely s mírou příslušnosti 1 absolutně náleží danému shluku, naopak 0 absolutně nenáleží.

S jistotou tedy můžeme říct, že shluk červené barvy, viz Obr. 29 (uprostřed) reprezentuje nejtmavší pixely v obraze, na rozdíl od shluku modré barvy, který vyobrazuje cévní řečiště oblasti krku a hlavy, vyznačující se pixely s nejvyšší intenzitou jasu.



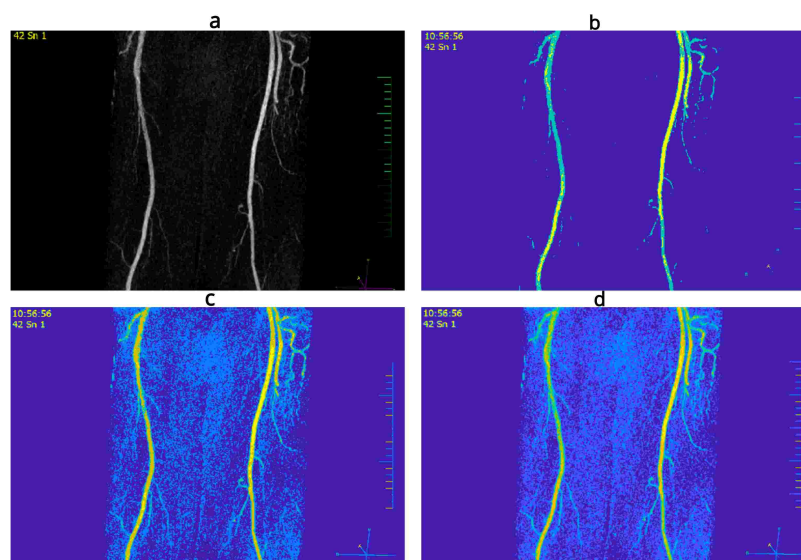
Obr. 30: Nativní snímek (vlevo), Segmentovaný obraz metodou fuzzy c-means (uprostřed), graf míry příslušnosti jednotlivých pixelů ke shlukům (vpravo)

Obr. 30 reprezentuje angiografický záznam cévního zásobení krku a hlavy pořízený magnetickou rezonancí. Na levé straně je vyobrazen nativní snímek v jednotlivých odstínech šedi. Prostřední snímek je po segmentaci metodou fuzzy c-means a koresponduje s jednotlivými grafy na pravé straně Obr. 30. Tyto grafy reprezentují míru příslušnosti jednotlivých pixelů ke shlukům. V tomto případě pro počet 5 shluků (nahore) a 6 shluků (dole) výše zmíněného Obr. 30.

7 Testování metod na reálných medicínských datech

Tato část práce je věnována testování na reálných medicínských datech a jejich následnému porovnání na základě implementace metod nehierarchické shlukové analýzy k-means a fuzzy c-means. Na následujících testovacích medicínských snímcích jsou vždy vyobrazeny na jedné straně nativní a na druhé segmentované snímky.

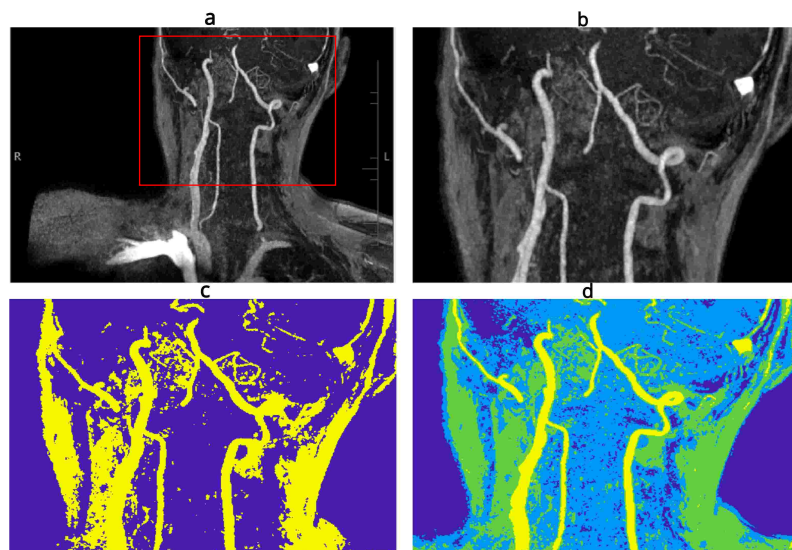
Nastavované počty segmentačních tříd jsou voleny v různých počtech, a to od 2 do 8. Na vybrané snímky byl aplikován aditivní šum určitého typu a intenzity. Na snímcích je možné rozeznat rozdíl segmentačních úrovní a parazitního šumu. Implementace byla provedena především na data pořízená za pomoci magnetické rezonance. U některých typů snímků jsou vybrány jejich preferované oblasti, které jsou poté separátně segmentovány. Ukázka testovaných datasetů za účelem následné evaluace je vyobrazena v příloze III.



Obr. 31: (a) – nativní snímek, (b) – segmentovaný snímek metodou k-means pro 3 shluky, (c) – segmentovaný snímek metodou k-means pro 5 shluků, (d) – segmentovaný snímek metodou k-means pro 8 shluků

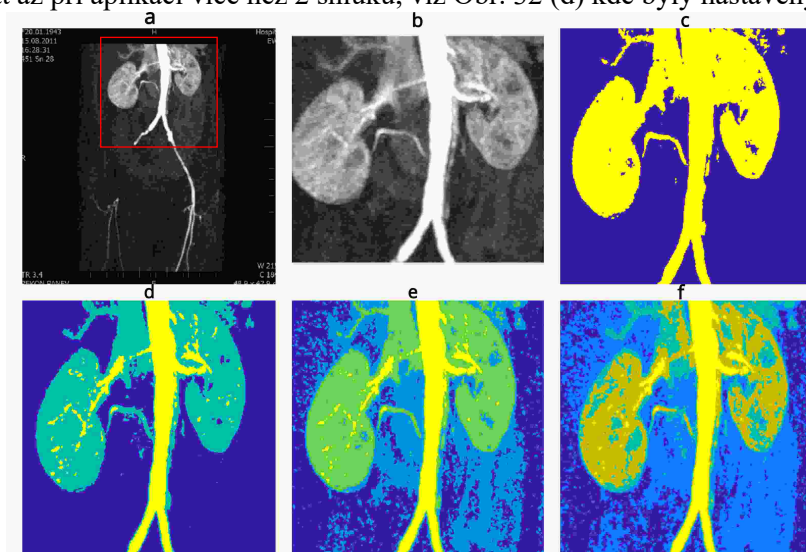
Na Obr. 31 (a) je vyobrazeno cévní zásobení dolních končetin, snímky pochází z magnetické rezonance a jedná se o angiografii. Pořízený nativní snímek na Obr. 31 (a) disponuje jasově odlišnými částmi, a to pouze v oblasti cév, kde byla provedena aplikace kontrastní látky.

Pozadí a okolí cév je nepotřebné, proto při nastavení 3 shluků dojde k jeho odfiltrování, viz Obr. 31 (b). U nastavení 5 a 8 shluků na Obr. 31 (c, d) je pozadí opět viditelné, ale zobrazení cév zůstává duplicitní, proto se zde jeví nastavení 3 shluků jako nejlepší řešení.



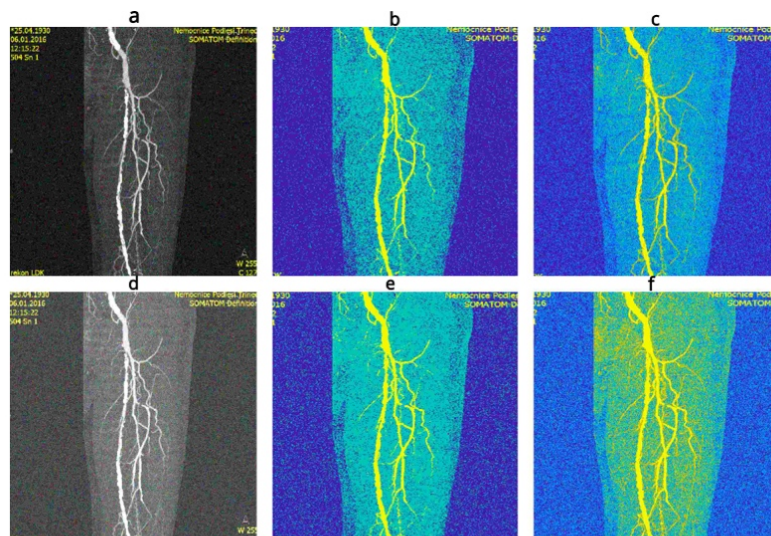
Obr. 32: (a) – nativní snímek, (b) – nativní snímek vybrané oblasti, (c) – segmentovaný snímek metodou k-means pro 2 shluky, (d) – segmentovaný snímek metodou k-means pro 4 shluky

Segmentace obrazu nemusí být účelně implementována na celý snímek, ale pouze na jeho důležité podoblasti. V případě Obr. 32 (a, b) byla vybrána část snímku zachycující cévní řečiště v oblasti krku. Je zřejmé, že při nastavení 2 shluků, jak ukazuje Obr. 32 (c), není metoda k-means schopna cévy rozeznat. Snímek je totiž rozdělen pouze na pozadí (jeden shluk) a tkáň (druhý shluk). Řečiště jsme schopni rozeznat až při aplikaci více než 2 shluků, viz Obr. 32 (d) kde byly nastaveny 4 shluky.



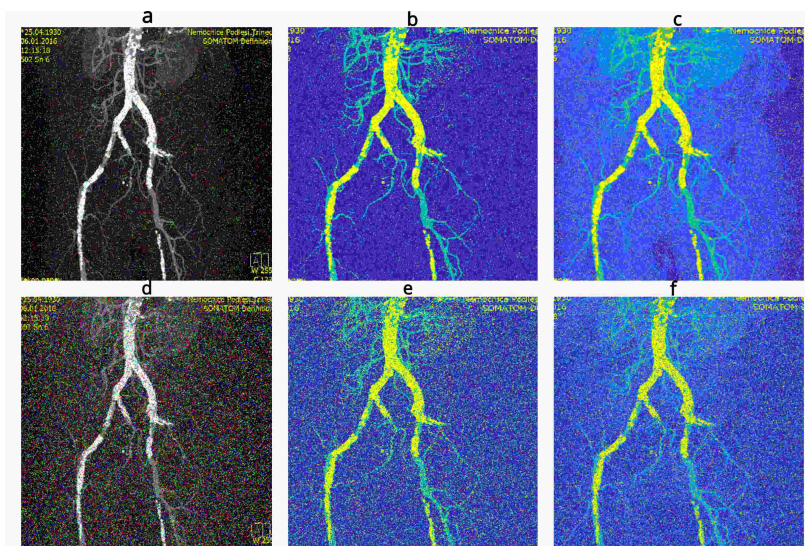
Obr. 33: (a) – nativní snímek, (b) – nativní snímek vybrané oblasti, (c) – segmentovaný snímek metodou fuzzy c-means pro 2 shluky, (d) – segmentovaný snímek metodou fuzzy c-means pro 3 shluky, (e) – segmentovaný snímek metodou fuzzy c-means pro 4 shluky, (f) – segmentovaný snímek metodou fuzzy c-means pro 5 shluků

Na Obr. 33 je vidět opět provedená selekce části snímku zachycující ledviny. Segmentace byla provedena za pomoci metody fuzzy c-means. Jednotlivé snímky od a do f zobrazují stupňující se nastavený počet shlukovacích tříd. Na Obr. 33 (c) jsou ledviny špatně rozpoznatelné. Nejlépe segmentovaný snímek je na Obr. 33 (d) při nastavení 3 shluků, neboť byly potlačeny parazitní části pozadí a zůstaly pouze důležité části snímku.



Obr. 34: (a) – implementace gaussovského šumu, ($\mu=0$, $\sigma=0,1$), (b) – segmentovaný snímek metodou *k-means* pro 3 shluky, (c) – segmentovaný snímek metodou *k-means* pro 5 shluků, (d) – implementace gaussovského šumu, ($\mu=0$, $\sigma=0,3$), (e) – segmentovaný snímek metodou *k-means* pro 3 shluky, (f) – segmentovaný snímek metodou *k-means* pro 5 shluků

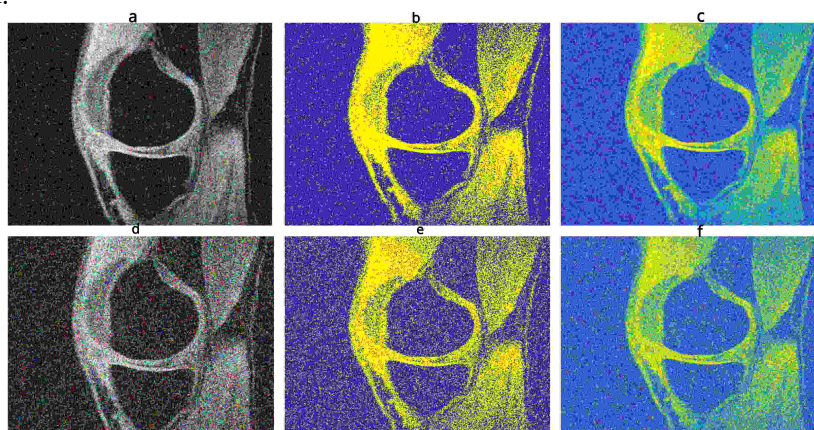
Obrazová data generovaná za pomoci zobrazovacích metod v medicíně jsou ve většině případů ovlivněna určitou mírou a typem šumu. Na Obr. 34 (a) je nastaven umělý aditivní šum gaussovského typu na hodnotu, ($\mu=0$, $\sigma=0,1$) a na Obr.34 (d) na hodnotu, ($\mu=0$, $\sigma=0,3$). Následně je implementována metoda *k-means*, a to s nastavením 3 shluků na Obr. 34 (b, e) a s nastavením 5 shluků na Obr. 34 (c, f). Snímek pochází opět z angiografického záznamu dolní končetiny při aplikaci kontrastní látky.



Obr. 35: (a) – implementace salt&pepper šumu, ($d=0,1$), (b) – segmentovaný snímek metodou *k-means* pro 3 shluky, (c) – segmentovaný snímek metodou *k-means* pro 8 shluků, (d) – implementace salt&pepper šumu, ($d=0,3$), (e) – segmentovaný snímek metodou *k-means* pro 3 shluky, (f) – segmentovaný snímek metodou *k-means* pro 8 shluků

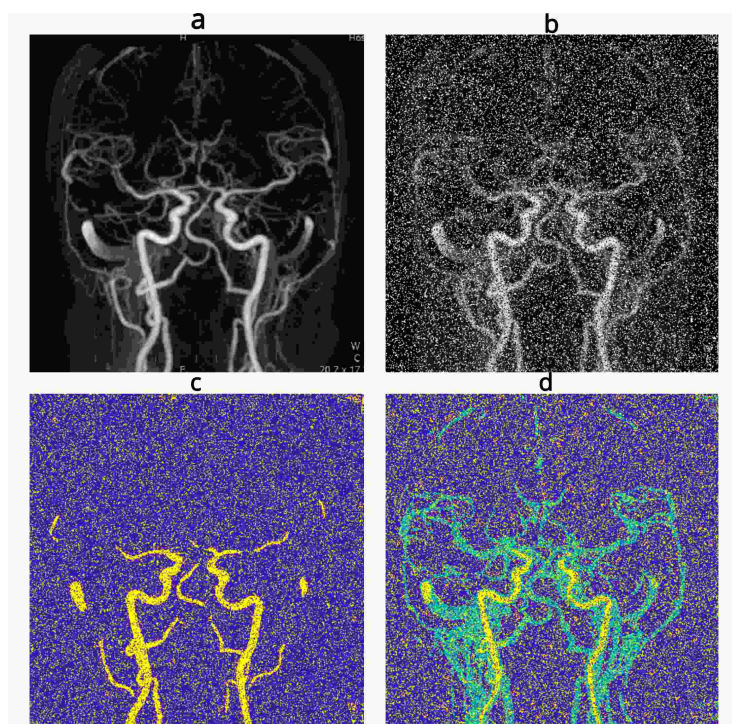
Druhým typem z implementovaných aditivních šumů je salt & pepper. Obraz pochází z angiografie dolních končetin a pánve, získané za pomoci magnetické rezonance. Na Obr. 35 (a) je denzita nastavena na 0,1 a 0,3 na Obr. 35 (d). Následně je obraz segmentován metodou *k-means* pro 3 shluky, jak ukazuje Obr. 35 (b, e) a pro 8 shluků, viz Obr.35 (c, f). V obou případech nastavení je

zřetelné, že k-means je schopna rozlišit a přenést do popředí jednotlivé typy tkání i za přítomnosti aditivního šumu.



Obr. 36: (a) – implementace localvar šumu, ($\mu=0$, $\sigma=0,1$), (b) – segmentovaný snímek metodou fuzzy c-means pro 2 shluky, (c) – segmentovaný snímek metodou fuzzy c-means pro 6 shluků, (d) – implementace localvar šumu, ($\mu=0$, $\sigma=0,3$), (e) – segmentovaný snímek metodou fuzzy c-means pro 2 shluky, (f) – segmentovaný snímek metodou fuzzy c-means pro 6 shluků

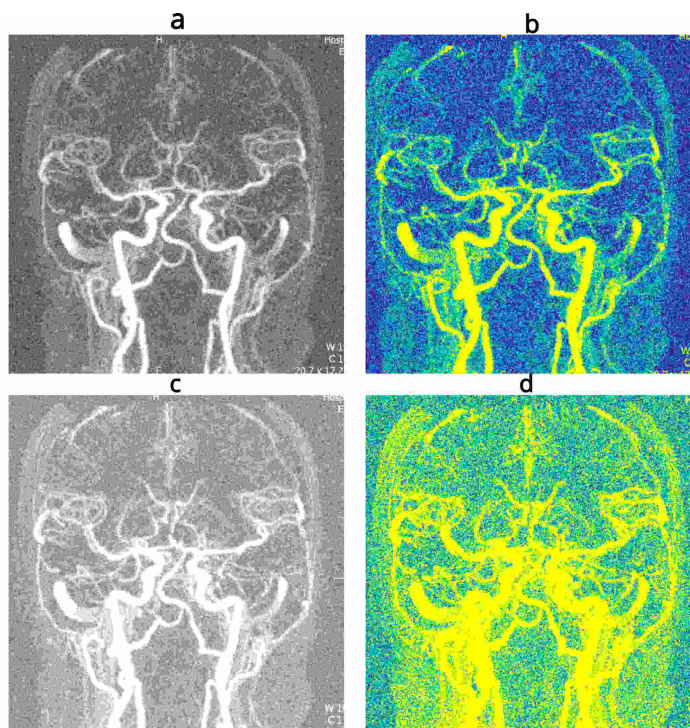
Na Obr. 36 (a) je vyobrazena artikulární chrupavka kolenního kloubu s nastavením localvar šumu s úrovní intenzity, ($\mu=0$, $\sigma=0,3$) a poté silným zašuměním s úrovní, ($\mu=0$, $\sigma=0,3$), viz Obr. 36 (d). Segmentace je provedena v počtu 2 shluků, viz Obr. 36 (b, e). Na snímku lze vidět, že metoda obraz úspěšně rozdělila na 2 shluky, a to na pozadí a snímání kolenní kloub, detaily snímku jsou potlačeny do pozadí. Na Obr 36 (c, f) jsou při nastavení 6 shluků patrné vyšší detaily i u jasově nevýrazných částí kloubu.



Obr. 37: (a) – nativní snímek, (b) – implementace salt&pepper šumu, ($d=0,3$), (c) – segmentovaný snímek metodou fuzzy c-means pro 2 shluky, (d) – segmentovaný snímek metodou fuzzy c-means pro 3 shluky

Na Obr. 37 (a) je vyobrazený nativní snímek cévního řečiště oblasti hlavy získaný za pomoci angiografie. Jasná a nejvíce ostrá část snímku je aplikovaná kontrastní látka kolující v cévách. Na Obr. 37 (b) je implementován šum salt & pepper s vysokou hustotou. Je patrné, že cévní řečiště zaniká spolu s pozadím. Na Obr. 37 (c, d) je již implementována metoda fuzzy c-means.

Při použití 2 shluků na Obr. 37 (c) je řečiště metodou úspěšně rozpoznáno a vyznačeno shlukem žluté barvy, vše ostatní bylo potlačeno do pozadí. Při nastavení pro 3 shluky se na Obr. 37 (d) vyobrazí i okolní tkáň reprezentované shlukem zelené barvy, avšak již dochází ke zkreslení krevního řečiště.



Obr. 38: (a) – implementace gaussovského šumu, ($\mu=0$, $\sigma=0,3$), (b) – segmentovaný snímek metodou fuzzy c-means pro 3 shluky, (c) – implementace gaussovského šumu ($\mu=0$, $\sigma=0,4$), (d) – segmentovaný snímek metodou fuzzy c-means pro 5 shluků

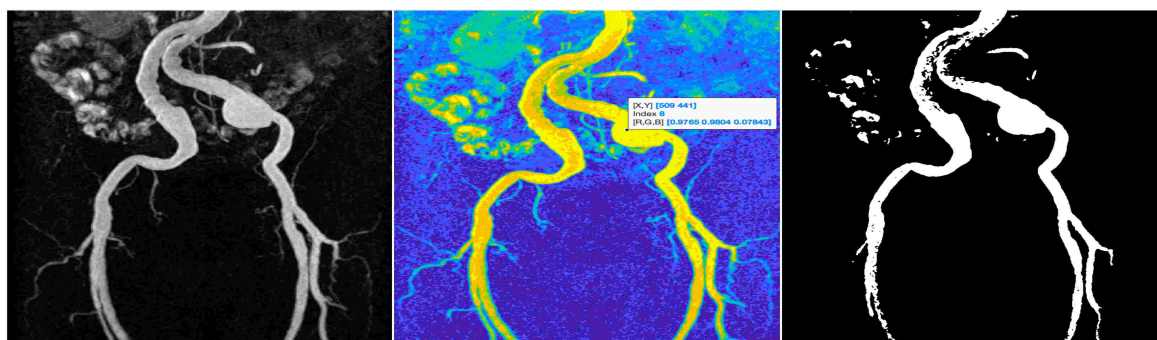
Následující Obr. 38 (a, c) ukazuje silně zašuměné snímky cévního řečiště v oblasti krku a hlavy. Na snímky byl aplikován gaussovský šum o intenzitě, ($\mu=0$, $\sigma=0,3$) a ($\mu=0$, $\sigma=0,4$), viz Obr.38 (b, d). Vyšší intenzita je zvolena za účelem stanovení hranic metody a jejího rozpoznání jednotlivých oblastí snímku. Použita je metoda fuzzy c-means, která je robustnější na data s intenzivním šumem.

Nastavení pro gaussovský šum, ($\mu=0$, $\sigma=0,3$) a 3 shluky metoda úspěšně segmentovala cévy, viz Obr. 38 (b). Na rozdíl od nastavení pro 5 shluků a intenzitu, ($\mu=0$, $\sigma=0,4$), dochází k velkému zkreslení a segmentace je v tomto případě neúspěšná.

8 Modelování tkání na základě shlukové analýzy

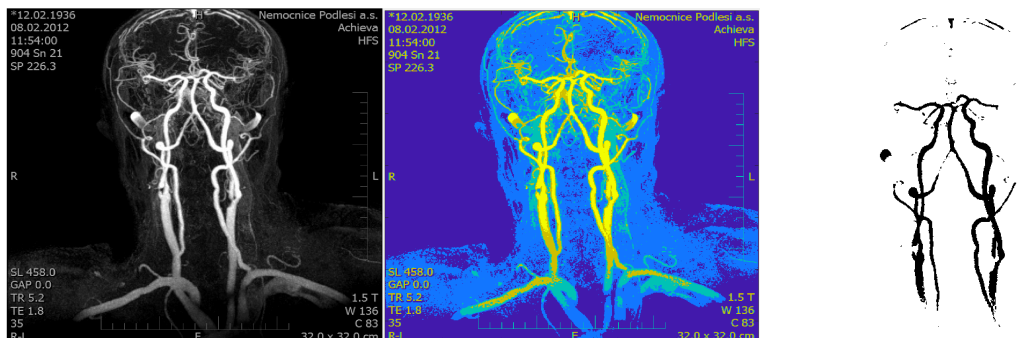
Modelování tkání je nezbytnou částí zpracování medicínských obrazů. Dochází zde díky matematickému popisu a následné aplikaci vhodného segmentačního algoritmu k odlišení důležitých objektů v obraze. Takovým objektem může být kupříkladu nádor mozku, který jsme díky těmto metodám schopni detekovat, ohraničit, či dokonce spočítat jeho charakteristické příznaky.

V případě Obr. 39 byl obraz segmentován metodou fuzzy c-means a byl rozdělen do 8 shluků zobrazených odlišnými barvami. Následně byl aktivován kurzor, kterým bylo zjištěno, že pixely vyobrazující cévní řečiště spadají dle indexační matice ke shluku s indexem 8 (pixely žluté barvy). Následujícím krokem byl obraz procházen a všechny indexy nerovnáající se shluku 8 byly vynulovány, tudíž se zobrazí jako pixely černé barvy.



Obr. 39: Vlevo nativní snímek, uprostřed segmentovaný snímek s nastavením 8 shluků, vpravo extrakce cévního zásobení dolních končetin

Na Obr. 40. byl obraz segmentován metodou k-means do 5 shluků. Při pohledu na obraz vpravo lze vidět vyextrahovaný cévní systém zásobení mozku. Cévy jsou na rozdíl od Obr. 39. zastoupeny pixely černé barvy, tudíž je zde docíleno opačného principu, a to vynulování shluku, který chceme zobrazit. V našem případě se jedná právě o cévní systém.



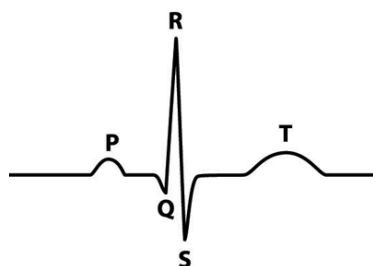
Obr. 40: Vlevo nativní snímek, uprostřed segmentovaný snímek s nastavením 5 shluků, vpravo extrakce cévního zásobení mozku

Cílem modelování tkání je převedení vstupních obrazových dat generovaných zobrazovacím systémem (MRI, CT) na matematickou informaci o daných objektech, které se v obraze nacházejí. Tyto objekty jsou ve většině případech tvořeny pixely s odlišnými vlastnostmi, díky tomu jsme schopni rozlišit vlastnost těchto pixelů a přiřadit je následovně k jednotlivým tkáním. Takto rozčleněné části obrazu je poté možné jako celky odfiltrovat a zobrazit si pouze ty, které jsou pro uživatele důležité.

9 Implementace metod shlukové analýzy na jednorozměrné signály

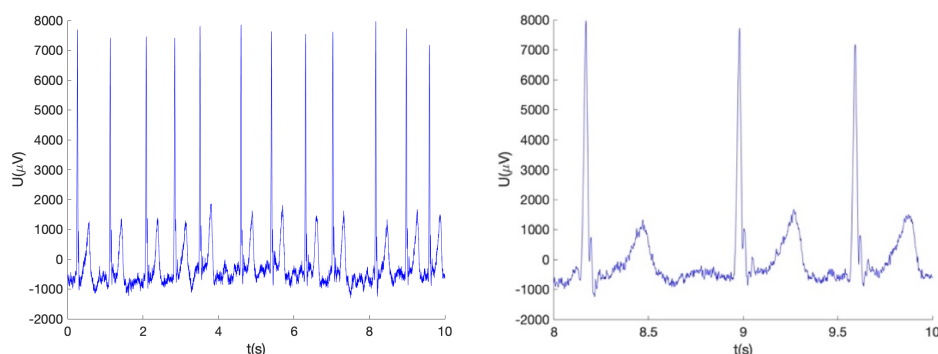
Kapitola se zabývá možnostmi implementace shlukové analýzy pro 1D signály s cílem dekompozice charakteristických segmentů těchto signálů.

Metody k-means a fuzzy c-means byly implementovány na EKG záznam. EKG je zařízení, které je schopné měřit elektrickou aktivitu myokardu srdce na povrchu těla pacienta, jelikož ta se šíří všemi směry do celého těla. Výstupem je EKG křivka, viz Obr. 41. Křivka se skládá z jednotlivých částí. Úsek P-Q reprezentuje depolarizaci síní, následuje QRS komplex, který reprezentuje stah svaloviny komor a vypuzení krve do oběhu. Vlna T vzniká při repolarizaci srdečních komor.



Obr. 41: Křivka EKG záznamu [24]

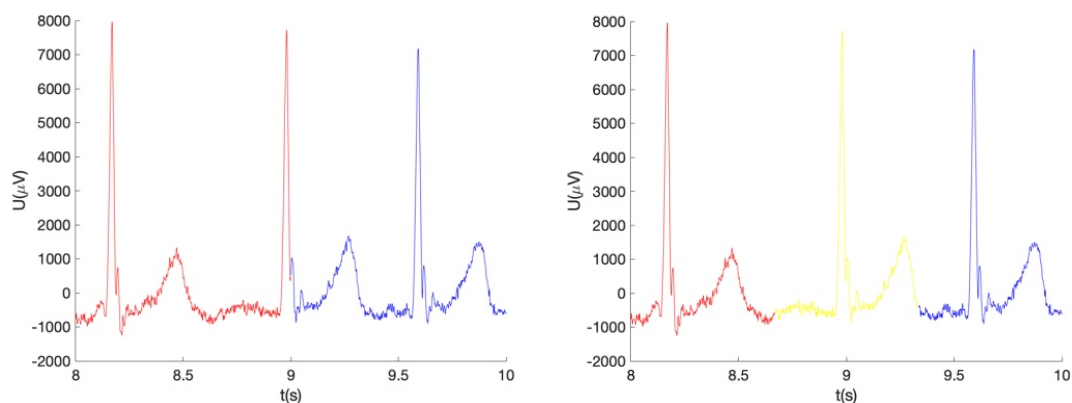
V první řadě byla vybrána fyziologická křivka, která je vyobrazena na Obr. 42 (vlevo). Následně byl extrahován vzorek křivky o délce 2 sekund. Křivka vykresluje 3 srdeční stahy, úkolem shlukové analýzy je v tomto případě rozčlenění křivky na 3 části. Tím dosáhneme oddělení jednotlivých srdečních stahů na samostatné celky.



Obr. 42: Vlevo fyziologická křivka EKG, vpravo extrakce části křivky

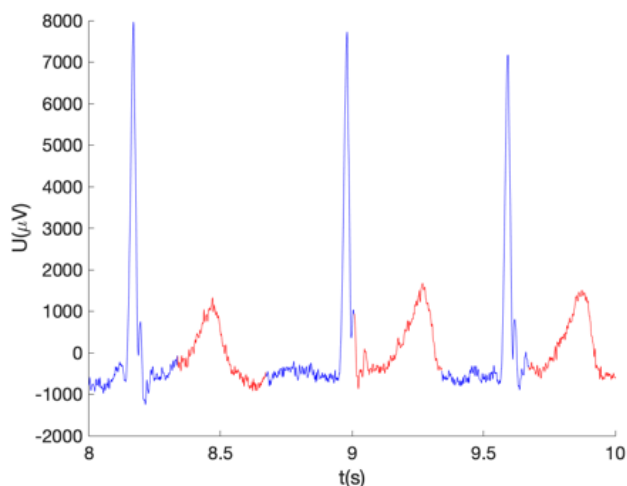
Následně byla implementována metoda k-means. V prvním případě byla nastavena pro 2 shluky, viz Obr. 43 (vlevo). Toto nastavení se jeví jako neefektivní řešení, neboť se křivka rozdělí pouze na shluk modré a červené barvy, tudíž na dvě části.

Úkolem segmentace křivky je klasifikace jednotlivých srdečních stahů. Toho je docíleno na Obr. 43 (vpravo). V tomto případě bylo správným řešením zvolení 3 shluků, to znamená, že je křivka rozdělena na jednotlivé části a každá zmíněná část reprezentuje jeden srdeční stah.

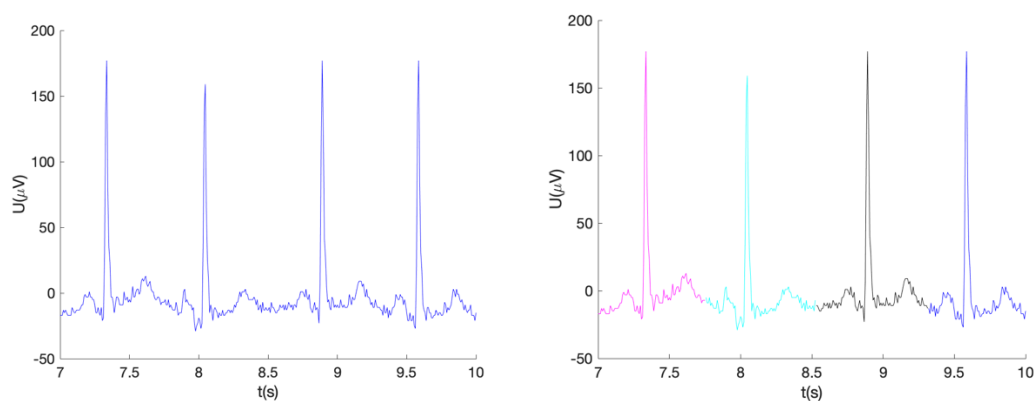


Obr. 43: Vlevo segmentace křivky metodou *k-means* s nastavením 2 shluků, vpravo segmentace křivky metodou *k-means* s nastavením 3 shluků

Při nastavení více shluků v případě Obr. 44 jsme schopni křivku rozčlenit na části menší než jeden srdeční stah. V tomto nastavení došlo k rozdělení na 3 shluky modré barvy reprezentující depolarizaci síní a komor a 3 shluky červené barvy detekující repolarizaci komor, tedy ST úsek. Tento úsek je důležitou částí křivky pro zjištění závažného infarktu myokardu, který se vyznačuje nadměrným stoupáním či klesáním právě ST úseku. Následující extrakce byla provedena na základě metody fuzzy *c-means*.



Obr. 44: Segmentace křivky metodou fuzzy *c-means* pro extrakci ST úseku



Obr. 45: Vlevo EKG záznam srdeční arytmie, vpravo segmentace křivky pro 4 shluky

Následující křivka na Obr. 45 (vlevo) zobrazuje nefyziologickou srdeční činnost. Úkolem segmentace na základě metody fuzzy c-means je v tomto případě opět oddělení jednotlivých stahů pro následnou možnou analýzu a vyhledání anomálií. Byly zvoleny 4 shluky, jak ukazuje Obr. 45 (vpravo), což bylo správným řešením.

10 Evaluace testovaných dat

V Tab. 3 jsou výsledné hodnoty jednotlivých korelačních koeficientů pro různá nastavení segmentačních metod. Tyto koeficienty jsou vždy zprůměrovány ze 30 zpracovávaných snímků a jednotlivých šumů. Každá hodnota tabulky je tedy průměrným zástupcem skupiny jednotných výsledných hodnot.

Tab. 3: Výsledné průměrné korelace metod pro jednotlivé šumy

	k-means			fuzzy c-means		
	3 shluky	5 shluků	8 shluků	3 shluky	5 shluků	8 shluků
Gaussovský šum	0,589	0,640	0,653	0,581	0,623	0,638
Salt & pepper šum	0,718	0,743	0,756	0,718	0,751	0,758
Speckle šum	0,885	0,931	0,941	0,898	0,934	0,947
Localvar šum	0,704	0,745	0,764	0,695	0,736	0,752

Následující Tab. 4 reprezentuje jednotlivé průměry výsledných hodnot, avšak střední kvadratické chyby jednotlivých snímků.

Tab. 4: Výsledné průměrné střední kvadratické chyby metod pro jednotlivé šumy

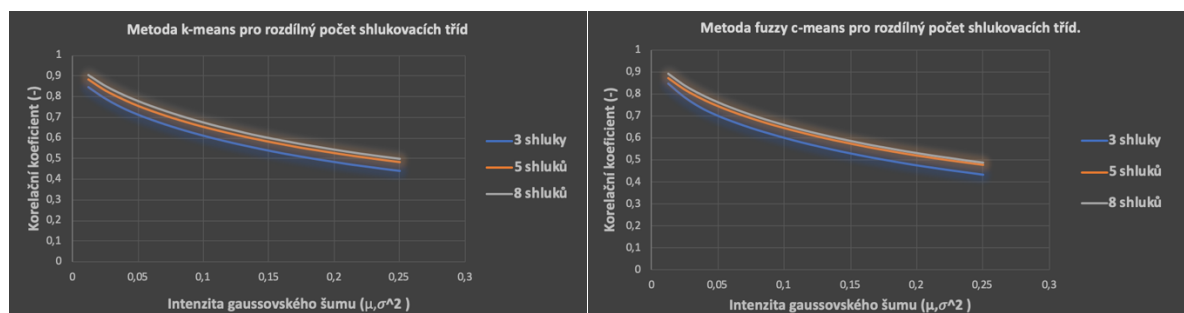
	k-means			fuzzy c-means		
	3 shluky	5 shluků	8 shluků	3 shluky	5 shluků	8 shluků
Gaussovský šum	0,418	1,123	2,909	0,430	1,233	3,487
Salt & pepper šum	0,283	0,685	1,583	0,300	0,706	1,693
Speckle šum	0,100	0,169	0,353	0,091	0,169	0,365
Localvar šum	0,266	0,705	1,771	0,281	0,791	2,212

Rozsáhlejší výsledky korelačních koeficientů a střední kvadratické chyby jsou zaznamenány v Tab. 1 a 2 přílohy IV. Současně jsou v této příloze uvedeny další grafické závislosti jednotlivých testování.

10.1 Analýza pro rozdílný počet shlukovacích tříd

Vstupem pro celkovou analýzu bylo u obou metod nutné zvolení počtu shluků. Tento parametr byl volen u obou metod stejně. Testování bylo provedeno s nastaveným gaussovským šumem pro následující grafy. Na levé straně Obr. 46 je vyobrazen graf pro metodu k-means a na pravé straně graf pro metodu fuzzy c-means. U obou metod jsou počty nastaveny na 3, 5 a 8 shluků.

Z grafů je patrné, že křivka pro 8 shluků vykazuje nejvyšší možné hodnoty korelačního koeficientu, a to s 10% rozdílem vůči křivce pro 3 nastavené shluky, z toho se dá logicky usuzovat, že větší počet shluků je více robustní na aditivní obrazový šum.

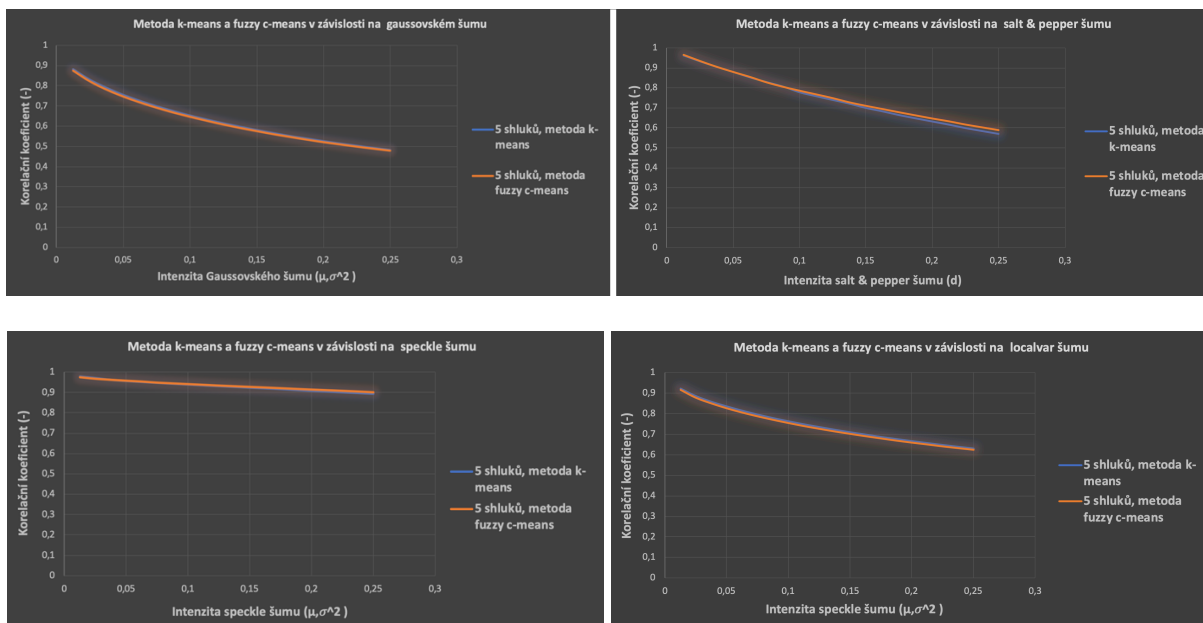


Obr. 46: Závislost počtu shluků na kvalitě segmentace obrazu

10.2 Analýza metody k-means a fuzzy c-means pro každý typ šumu

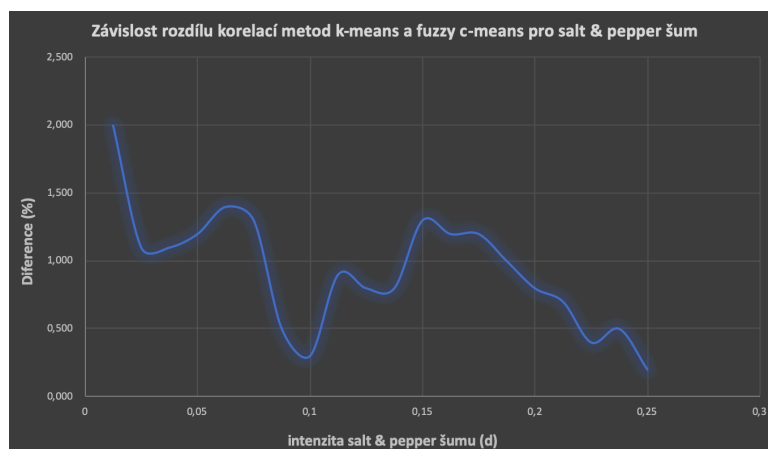
V této analýze bylo úkolem porovnat rozdílné výsledky mezi metodou k-means a fuzzy c-means pro každý z aplikovaných šumů. V jednotlivých grafech na Obr. 47 lze vidět křivky reprezentující korelační koeficienty metod v závislosti na intenzitě šumu.

Ačkoliv je metoda fuzzy c-means robustnější a její výpočet je matematicky náročnější, v této analýze jsou výsledné grafy pro obě metody téměř duplicitní. Co se týče lepší úspěšnosti jedné metody nad druhou, pro žádný ze šumů zde není patrný žádný markantní rozdíl. Detailnější rozbor difference lze vidět na Obr. 48.



Obr. 47: Srovnání metod k-means a fuzzy c-means na jednotlivých typech šumu

Na následujícím Obr. 48 je graficky vyobrazena difference mezi metodou k-means a fuzzy c-means s nastaveným šumem typu salt & pepper. Detailně je zde vidět procentuální rozdíl mezi oběma metodami, ačkoliv je minimální. Největší difference je v počátku při nižší intenzitě šumu, následně jsou výsledky obou metod s minimálním procentuálním rozdílem, a to zhruba do 1 %.

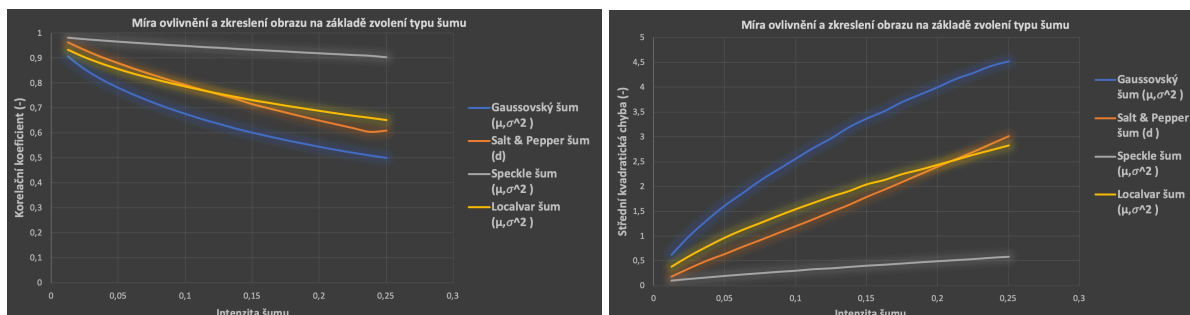


Obr. 48: Graf míry difference mezi metodami k-means a fuzzy c-means

10.3 Analýza míry zkreslení obrazu na základě zvolení typu šumu

Výsledné grafy reprezentují všechny druhy implementovaných šumů v závislosti na jejich intenzitě. Úkolem této analýzy bylo zjistit, který ze šumů nejagresivněji ovlivní medicínské obrazy.

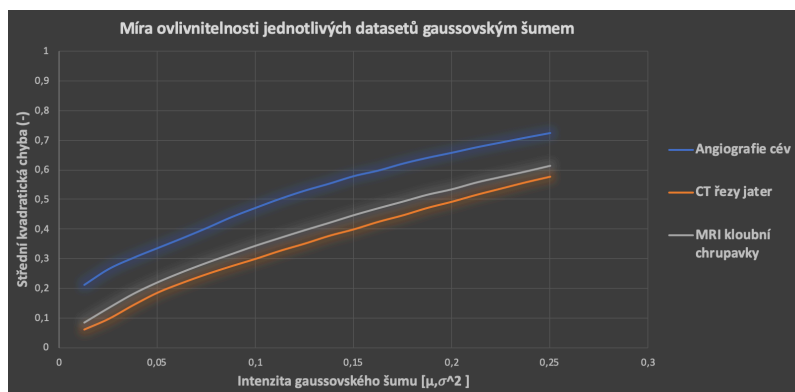
Nejméně ovlivňující byl vyhodnocen šum typu speckle, který v rámci korelačního koeficientu vykazuje pokles kvality obrazu na celkové intenzitě pouze 10 %. Je to dáno rozložením jeho šumu, které je situováno především na jasově intenzivnější oblasti obrazu. Nejvíce ovlivňující je gaussovský šum, který vykazuje zhruba 40% zkreslení obrazu. Důvodem je, že tento šum postihuje všechny pixely v obraze nezávisle na jejich jasové intenzitě. Šumy typu salt & pepper a localvar vykazují zkreslení přibližně 30 %.



Obr. 49: Grafy míry ovlivnění a zkreslení obrazu na základě zvolení typu šumu

10.4 Rozdíl ovlivnění obrazu šumem v závislosti na typu datasetu

Zde byla analyzována rozdílnost střední kvadratické chyby v závislosti na intenzitě gaussovského šumu pro každý z datasetů. Jako nejméně zasažené obrazy šumem vychází CT řezy jater a hned za nimi MRI kloubních chrupavek. Důvodem je především podobnost jednotlivých obrazů v datasetu na rozdíl od angiografie cév, kde se typ obrazů výrazně liší, a to především v úhlu a snímané části jednotlivých cévních partií. Výsledek angiografie cév je tudíž zhruba s 15% poklesem oproti ostatním datasetům.

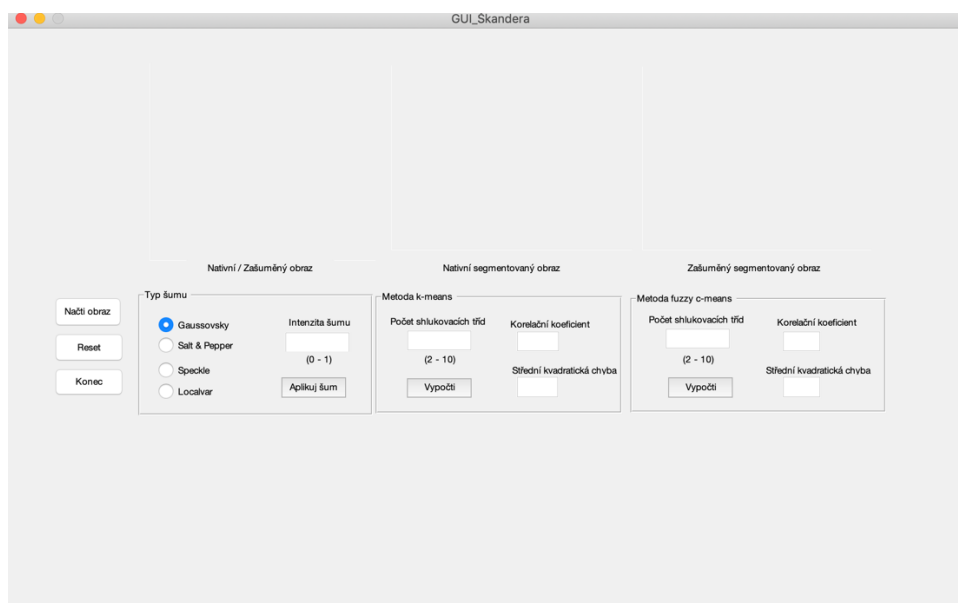


Obr. 50: Srovnání míry ovlivnitelnosti gaussovským šumem dle jednotlivých datasetů

11 Návrh graficko-uživatelského prostředí

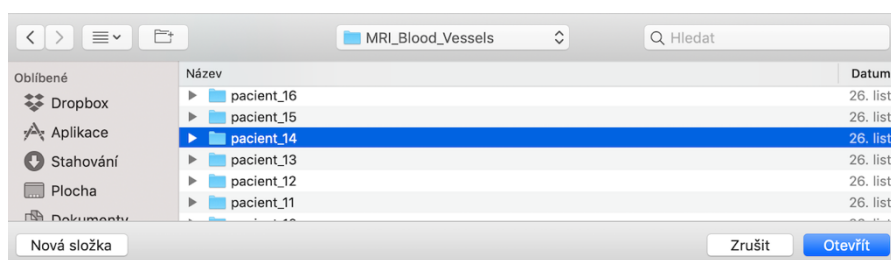
Na základě výše zpracovaných a implementovaných metod bylo vytvořeno v MATLAB graficko-uživatelské prostředí. V následující kapitole se nachází obrázky popisující jeho funkčnost.

Aplikace je členěna na 3 části. První z nich je na Obr. 51 (vlevo), zde probíhá načtení obrazu a implementace aditivního šumu. Následně má možnost uživatel volit mezi metodou k-means, která je nastavována ve střední části Obr. 51 a fuzzy c-means nastavována vpravo Obr. 51. U každé z metod je také zakomponován výpočet jednotlivého korelačního koeficientu a střední kvadratické chyby vůči nativnímu segmentovanému obrazu.



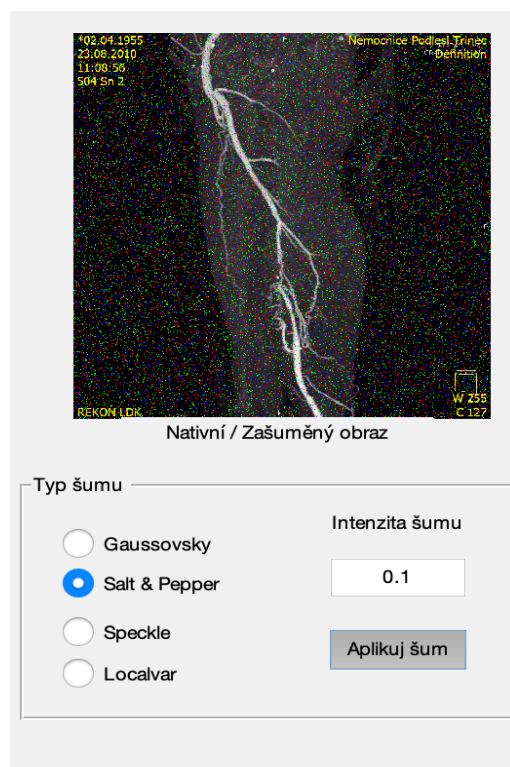
Obr. 51: Graficko-uživatelské prostředí

V prvním kroku je vybrán obraz z určitého datasetu. Za pomoci ovládacího prvku *Načti obraz* je vyvoláno dialogové okno a uživatel si vybírá jím zvolený obraz. Nahrání probíhá do pracovní plochy uživatelského prostředí. Součástí prostředí jsou také ovládací prvky *Reset*, pro uvedení aplikace do původního stavu a *Konec*, pro ukončení aplikace.



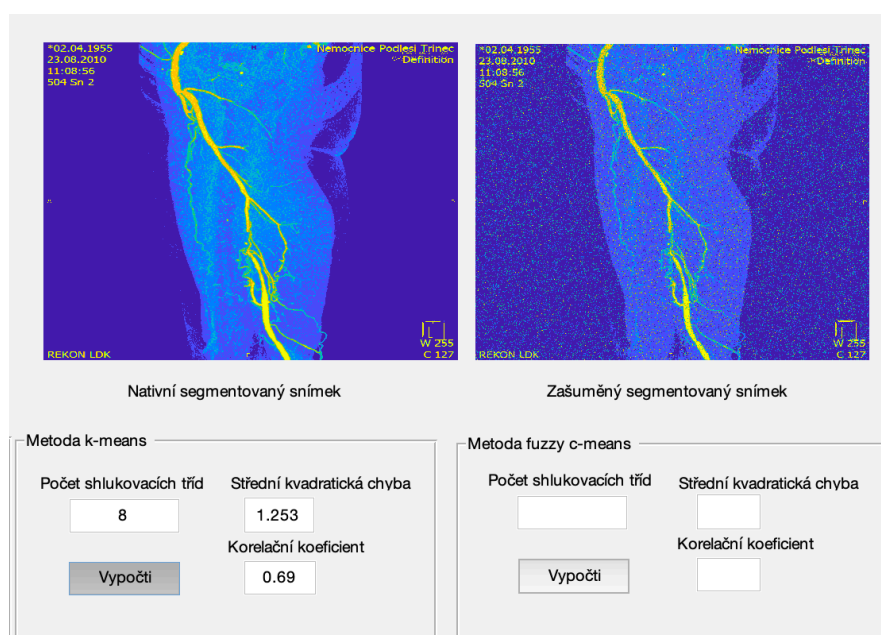
Obr. 52: Načtení obrazu do graficko-uživatelského prostředí

V druhém kroku je aplikován šum, kde za pomoci přepínače volíme jeden z typů šumu a jeho intenzitu. Intenzitu je možno volit zapsáním vybrané hodnoty do textového pole s názvem *Intenzita šumu*. Po stisknutí ovládacího prvku *Aplikuj šum* se na místo nativního snímku vyobrazí zašuměný obraz, viz Obr. 53.

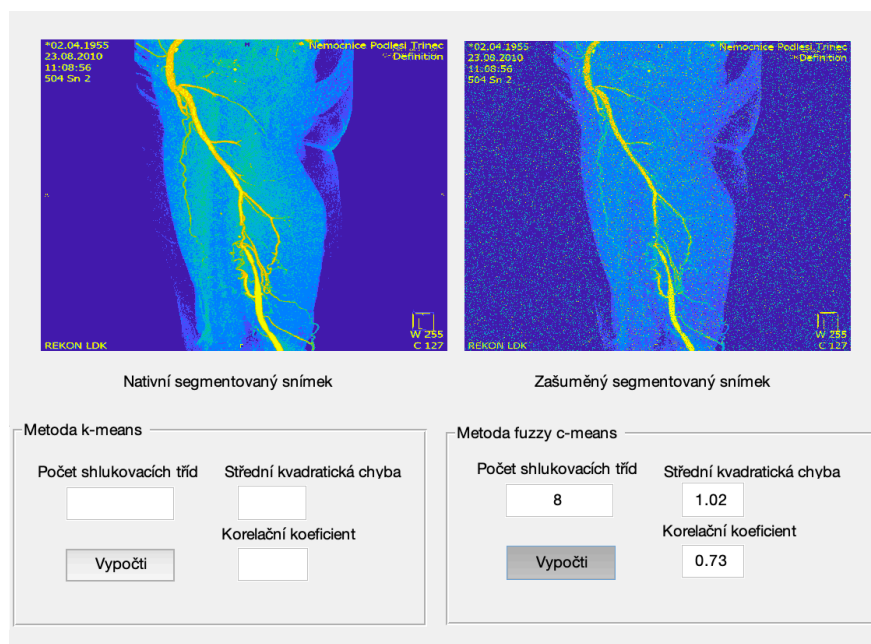


Obr. 53: Aplikace šum

V posledním kroku volíme ze dvou možností, mezi metodu k-means a fuzzy c-means. Obě metody mají v grafickém prostředí přidělené vlastní nastavení počtu shluků, textová pole pro zobrazení výsledku a ovládací prvek. Při stisknutí ovládacího prvku *Vypočti* je v první řadě vykreslen nativní segmentovaný obraz, poté zašuměný segmentovaný obraz a následovně je vypočten automaticky i korelační koeficient a střední kvadratická chyba. Tyto hodnoty jsou vypsány do příslušných textových polí.



Obr. 54: Volba metody k-means



Obr. 55: Volba metody fuzzy c-means

12 Závěr

Úkolem bakalářské práce bylo navrhnutí a design metod shlukové analýzy pro 1D a 2D medicínské signály. V prvním kroku byly shromážděny nezbytné informace o problematice učení bez učitele a klasifikaci biomedicínských dat, které jsou sepsány v teoretické části práce. Následně byly vybrány metody k-means a fuzzy c-means spadající do nehierarchické shlukové analýzy. Práce je z větší části koncipovaná na zpracování obrazových dat. Zpracováno bylo také testování na 1D signálech, a to analýza pro elektrokardiografickou křivku.

Hlavní implementace na obrazová data proběhla na vybrané 3 datasety získané především z magnetické rezonance a výpočetní tomografie. Na všechny obrazy byl aplikován aditivní šum. Následovalo testování obou metod při nastavení počtu prvotních shluků 3, 5 a 8. Každý jednotlivý zašuměný a nativní segmentovaný obraz byly následně mezi sebou porovnány. Porovnání bylo realizováno za pomoci korelačního koeficientu a střední kvadratické chyby. Všechny výsledné hodnoty se následně zprůměrovaly pro každý obraz a určitou hodnotu intenzity.

Závěrečná část je věnována zhodnocení testování. Hlavním úkolem bylo srovnání obou metod. Ačkoliv je metoda fuzzy c-means robustnější než k-means, v celkovém výsledku si jsou metody u všech druhů šumu výkonově téměř rovny. Odchylka se pohybuje v jednotkách procent. U analýzy zabývající se mírou ovlivnění obrazu dle typu šumu největší ovlivnění nastává při aplikaci gaussovského šumu, dochází až k 40% zkreslení vůči nativnímu obrazu. Nejmenší zkreslení, 10 %, bylo zaznamenáno u speckle šumu. U salt & pepper a localvar šumu bylo zkreslení do 30 %. Analýza pro zvolení počtu shluků vykazovala nejlepší výsledky u 8 shluků, avšak také delší výpočetní čas. Při srovnání s nastavením shluků 3 byl zde zhruba 15 % rozdíl. Z analýzy tedy vyplývá že vyšší počet shluků je více robustní na aditivní obrazový šum. Nejvíce ovlivněn šumem byl dataset s angiografickými snímky pořízen magnetickou rezonancí, důvodem je snímání různých částí těla pacienta, na rozdíl od obrazů kolenních chrupavek a CT řezů jater, kde byly obrazy získány z větší části pod stejným úhlem.

Testování proběhlo na jednorozměrných a dvojrozměrných medicínských datech, a to na základě implementace dvou zvolených metod. Do budoucna by se práce mohla zabírat trojrozměrnými obrazovými daty, extrakcí tkání v prostoru, měřením velikosti objektů v obraze a celkovým srovnáním pro tyto metody, které je možno také nadále rozšířit a optimalizovat.

Zdroje

- [1] GARRETA, Raul a Guillermo MONCECCHI. Learning scikit-learn: Machine Learning in Python. Birmingham: Packt Publishing Limited, 2013, 118 s. ISBN 978-1-78328-193-0.
- [2] MITCHELL, Tom M. Machine Learning. New York: McGraw-Hill, c1997. ISBN 978-0070428072.
- [3] HARUŠTIAKOVÁ, Danka. Vícerozměrné statistické metody v biologii. Brno: Akademické nakladatelství CERM, 2012. ISBN 978-80-7204-791-8.
- [4] HOLČÍK, Jiří. Analýza a klasifikace dat. Brno: Akademické nakladatelství CERM, 2012. ISBN 978-80-7204-793-2.
- [5] Marhold, K., Suda, J. Statistické zpracování mnohorozměrných dat v taxonomii (Fenologické metody). Učební texty Univerzity Karlovy v Praze. Univerzita Karlova v Praze, Nakladatelství Karolinum. (2002)
- [6] KLÍMEK, Petr. *Shlukovací metody v data miningu: E+M Ekonomie a Management* [online]. 2008, 120-127 [cit. 2018-11-19]. ISSN 1212-3609. Dostupné z: <http://www.ekonomie-management.cz/archiv/vyhledavani/detail/608-shlukovaci-metody-v-data-miningu/>.
- [7] STEINBACH, Michael, George KARYPIS a Vipin KUMAR. A Comparison of Document Clustering Techniques.[online]. University of Minnesota. Dostupné z: <http://glaros.dtc.umn.edu/gkhome/fetch/papers/docclusterKDDTMW00.pdf>.
- [8] GHUMAN, Sukhdev Singh. Clustering Techniques. *International Journal of Computer Science and Mobile Computing*[online]. Indie, 2016, 5.5.2016, 524-530 [cit. 2018-11-22]. ISSN 2320–088X.
- [9] RAMESH, B. a K. NANDHINI. Clustering Algorithms. *International Journal of Computer Sciences and Engineering*[online]. 2017, 30.10.2017, (5), 302-306 [cit. 2018-11-22]. ISSN 2347-2693.
- [10] MELOUN, Milan a Jiří MILITKÝ. *Statistická analýza experimentálních dat*. Vyd. 2., upr. a rozš. Praha: Academia, 2004. ISBN 80-200-1254-0.
- [11] ŘEZANKOVÁ, Hana, Dušan HÚSEK a Václav SNÁŠEL. *Shluková analýza dat*. Praha: Professional Publishing, 2007. ISBN 978-80-86946-26-9.
- [12] ABONYI, Janos a Balázs FEIL. *Cluster analysis for data mining and system identification*. Boston: Birkhäuser, c2007. ISBN 978-3764379872.

- [13] MELOUN, Milan a Jiří MILITKÝ. *Statistická analýza experimentálních dat*. Vyd. 2., upr. a rozš. Praha: Academia, 2004. ISBN 8020012540.
- [14] Holčík J, Komenda M (eds.) a kol. *Matematická biologie: e-learningová učebnice* [online]. 1. vydání. Masarykova univerzita [2015]. Dostupný z WWW: <http://portal.matematickabiologie.cz>. ISBN 978-80-210-8095-9.
- [15] WALNER, Jake. *Clustering RNAseq data using fuzzy c-means clustering* [online]. 31.7.2018 [cit. 2018-12-19]. Dostupné z: <https://2-bitbio.com/post/clustering-rnaseq-data-using-fuzzy-c-means-clustering/>
- [16] KAUFMAN, Leonard a Peter J. ROUSSEEUW. *Finding groups in data: an introduction to cluster analysis*. New York: Wiley, c1990. ISBN 9780471878766.
- [17] KUČERA, Jiří. Shluková analýza [online]. 2009 [cit. 2018-02-12]. Dostupné z: <http://aplikacergsg.sci.muni.cz/teorie/doprava/vicerozmerne-statisticke-metody-2>
- [18] LEISCH, Friedrich. Bagged Clustering. *Working Paper Series*. Augasse 2–6, 1090 Wien, Austria, 1999, (51).
- [19] JIRKOVSKÝ, Jaroslav. Strojové učení a prostředí MATLAB. In: *Humusoft* [online]. 2018 [cit. 2018-12-20]. Dostupné z: <https://www.humusoft.cz/blog/20180817-strojove-uceni/>
- [20] BARBU, Tudor. Variational Image Denoising Approach with Diffusion Porous Media Flow. *Abstract and Applied Analysis* [online]. 2013, 1-8 [cit. 2019-03-08]. DOI: 10.1155/2013/856876. ISSN 1085-3375. Dostupné z: <http://www.hindawi.com/journals/aaa/2013/856876/>
- [21] TOH, K.K.V. a N.A.M. ISA. Noise Adaptive Fuzzy Switching Median Filter for Salt-and-Pepper Noise Reduction. *IEEE Signal Processing Letters* [online]. 2010, 17(3), 281-284 [cit. 2019-03-08]. DOI: 10.1109/LSP.2009.2038769. ISSN 1070-9908. Dostupné z: <http://ieeexplore.ieee.org/document/5356178/>
- [22] HUA, Cheng a Tian JINWEN. Speckle Reduction of Synthetic Aperture Radar Images Based on Fuzzy Logic. In: *2009 First International Workshop on Education Technology and Computer Science* [online]. IEEE, 2009, 2009, s. 933-937 [cit. 2019-03-08]. DOI: 10.1109/ETCS.2009.212. ISBN 978-0-7695-3557-9. Dostupné z: <http://ieeexplore.ieee.org/document/4958916/>
- [23] ZAEFFERER, Martin. *Optimization and Empirical Analysis of an Event Detection Software for Water Quality Monitoring* [online]. Kolín, 2012 [cit. 2019-04-05]. Dostupné z: https://www.researchgate.net/publication/235710073_Optimization_and_Empirical_Analysis_of_an_Event_Detection_Software_for_Water_Quality_Monitoring
- [24] EKG.Academy: EKG Waveform Lesson. In: *EKG.Academy* [online]. [cit. 2019-04-21]. Dostupné z: <https://ekg.academy/ekg-waveform-lesson>.

Seznam příloh

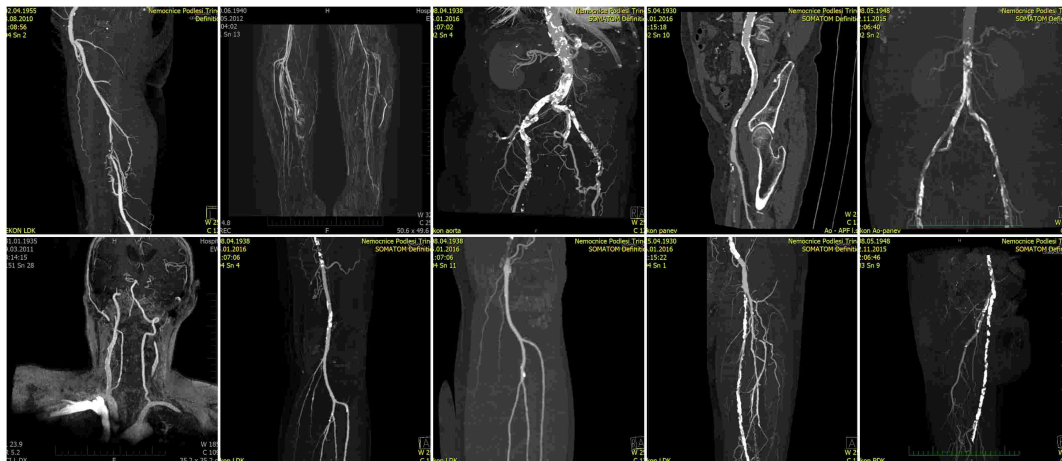
Příloha I.	Originální data
Příloha II.	Zašuměná data
Příloha III.	Výsledky metod k-means a fuzzy c-means
Příloha IV.	Grafy a tabulky
Příloha V.	Příloha na CD

I. Originální data

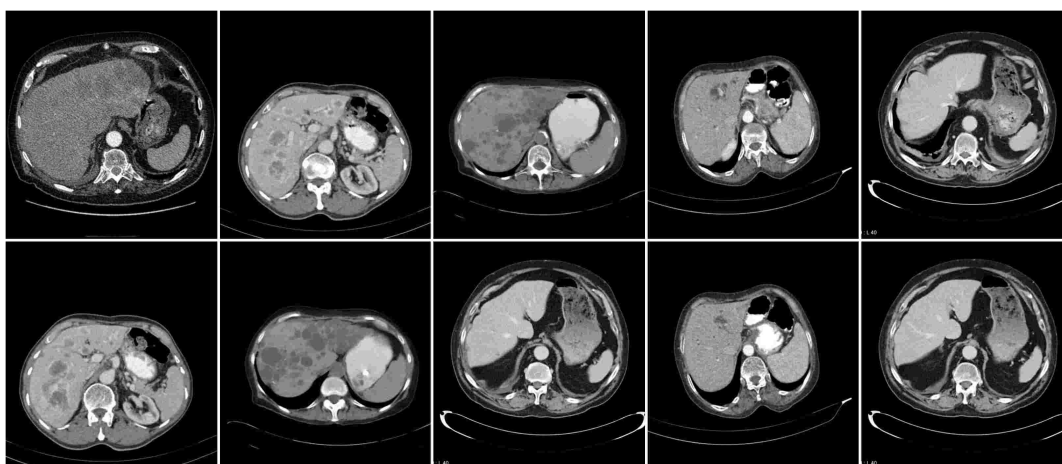
Obr.1: Dataset 1, cévní angiografie

Obr.2: Dataset 2, CT řezy jater

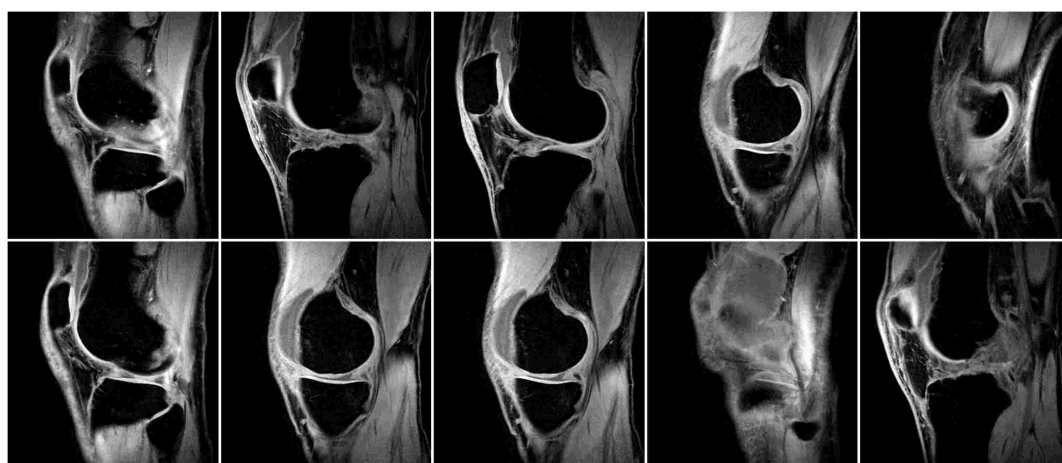
Obr.3: Dataset 3, MRI kolenní chrupavky



Obr. 1: dataset 1, cévní angiografie



Obr. 2: dataset 2, CT řezy jater



Obr. 3: dataset 3, MRI kolenní chrupavky

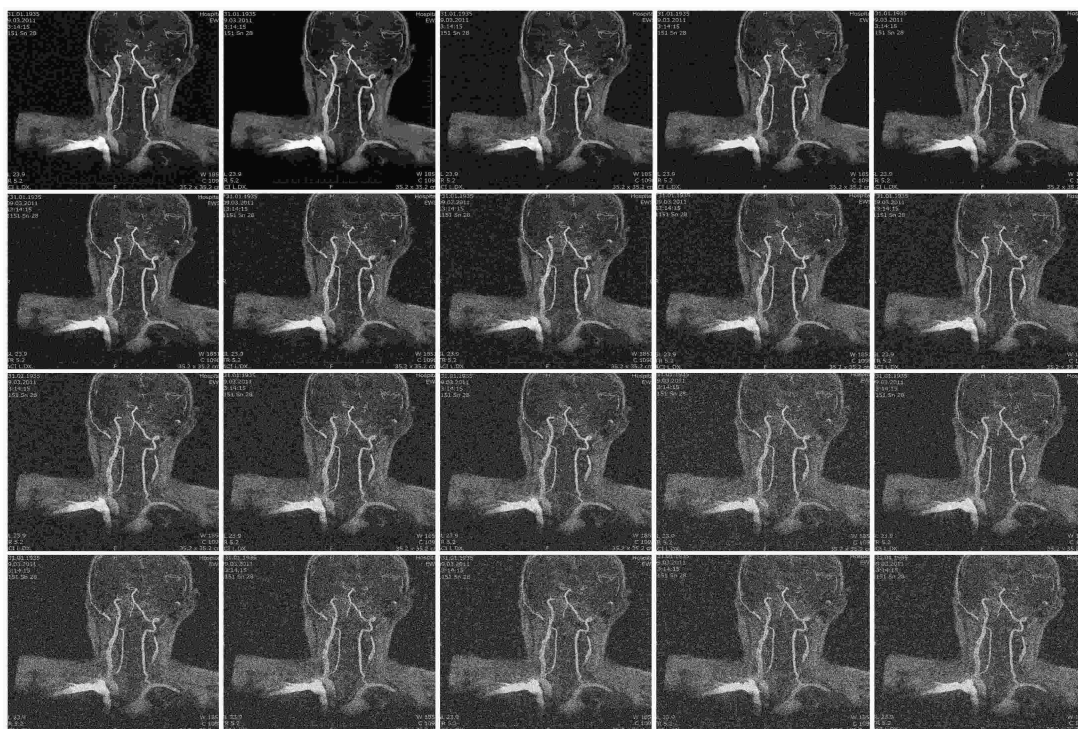
II. Zašuměná data

Obr.1: Implementace 20 úrovní gaussovského šumu, ($\mu=0$, $\sigma=0 - 0,25$) na datset č. 1

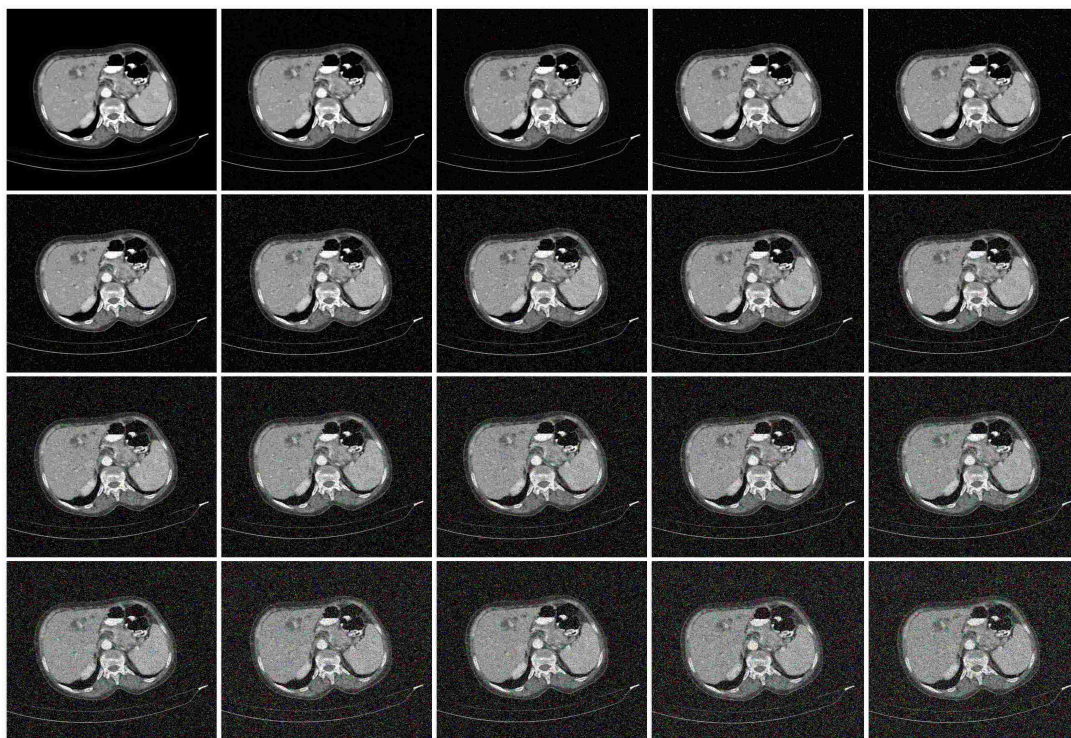
Obr.2: Implementace 20 úrovní salt & pepper šumu, ($d=0 - 0,25$) na dataset č. 2

Obr.3: Implementace 20 úrovní speckle šumu, ($\mu=0$, $\sigma=0 - 0,25$) na dataset č. 3

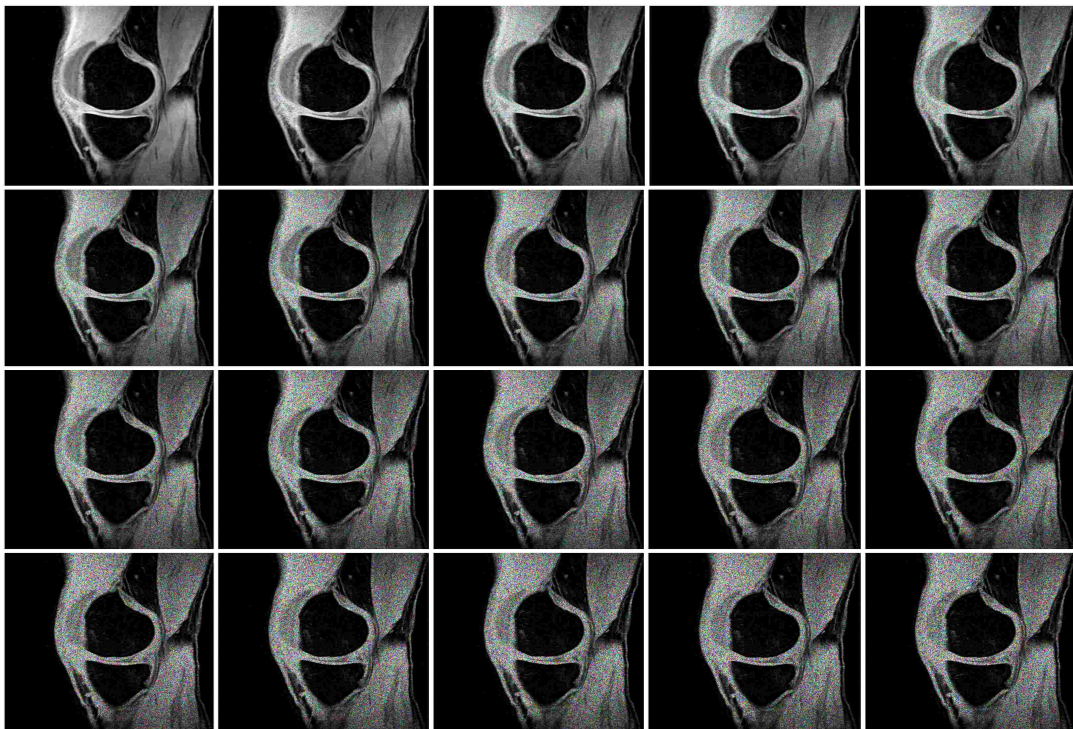
Obr.4: Implementace 20 úrovní localvar šumu, ($\mu=0$, $\sigma=0 - 0,25$) na dataset č. 1



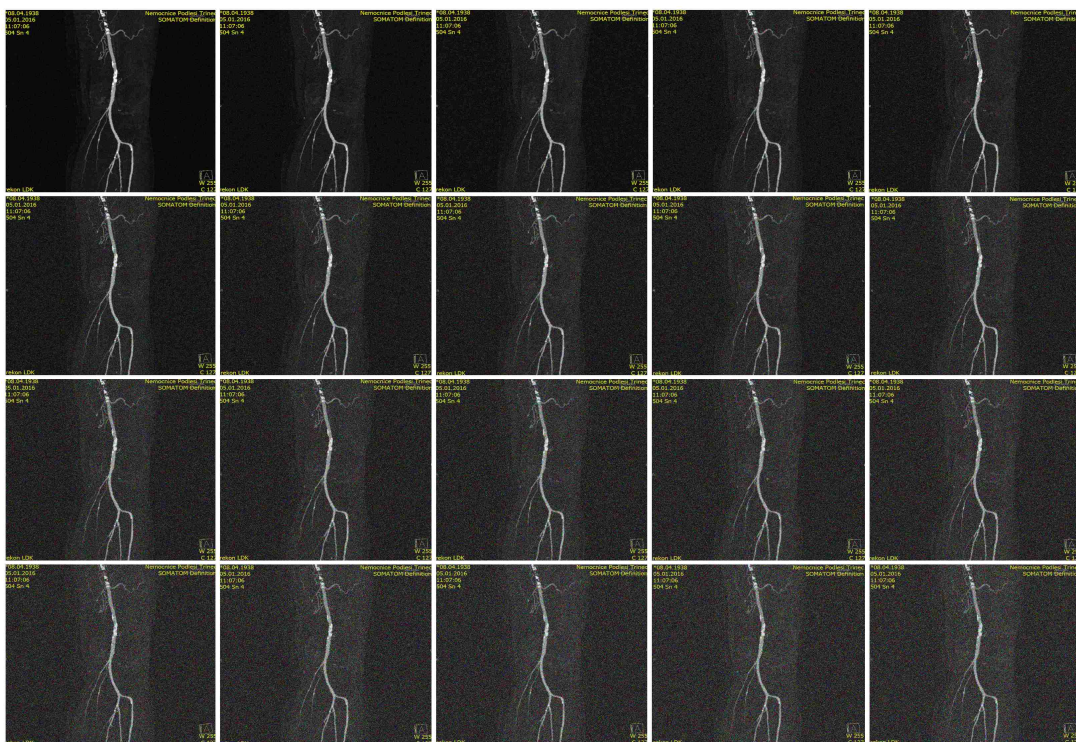
Obr. 1: Implementace 20 úrovní gaussovského šumu, ($\mu=0$, $\sigma=0 - 0,25$) na datset č. 1



Obr. 2: Implementace 20 úrovní salt & pepper šumu ($d=0 - 0,25$) na dataset č. 2



Obr. 3: Implementace 20 úrovní speckle šumu ($\mu=0$, $\sigma=0 - 0,25$) na dataset č. 3



Obr. 4: Implementace 20 úrovní localvar šumu ($\mu=0$, $\sigma=0 - 0,25$) na dataset č. 1

III. Výsledky metod k-means a fuzzy c-means

Obr.1: K-means pro 3 shluky implementována na obrazy s gaussovským šumem, ($\mu=0$, $\sigma=0 - 0,25$)

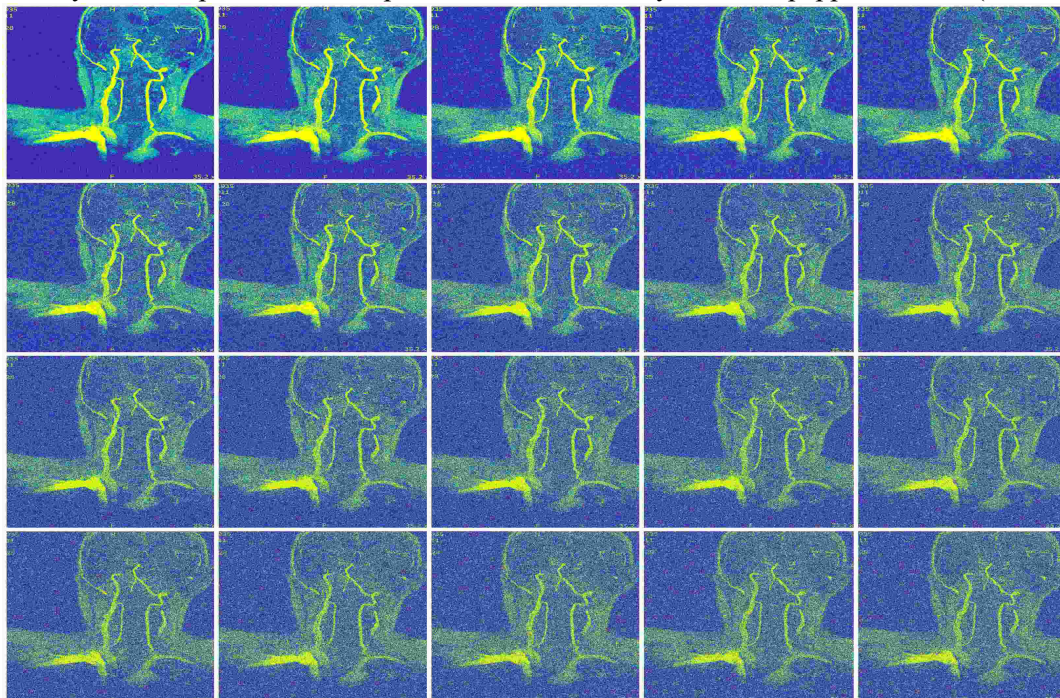
Obr.2: K-means pro 8 shluků implementována na obrazy s gaussovským šumem, ($\mu=0$, $\sigma=0 - 0,25$)

Obr.3: K-means pro 3 shluky implementována na obrazy se speckle šumem, ($\mu=0$, $\sigma=0 - 0,25$)

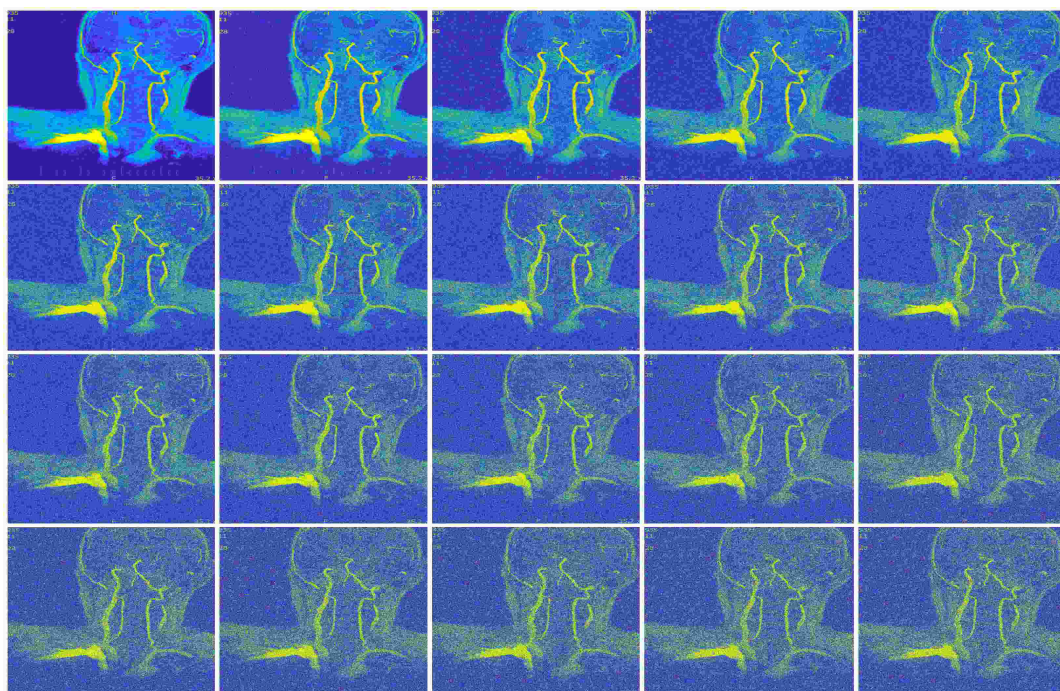
Obr.4: Fuzzy c-means pro 8 shluků implementována na obrazy se speckle šumem, ($\mu=0$, $\sigma=0 - 0,25$)

Obr.5: Fuzzy c-means pro 5 shluků implementována na obrazy se salt & pepper šumem, ($d=0 - 0,25$)

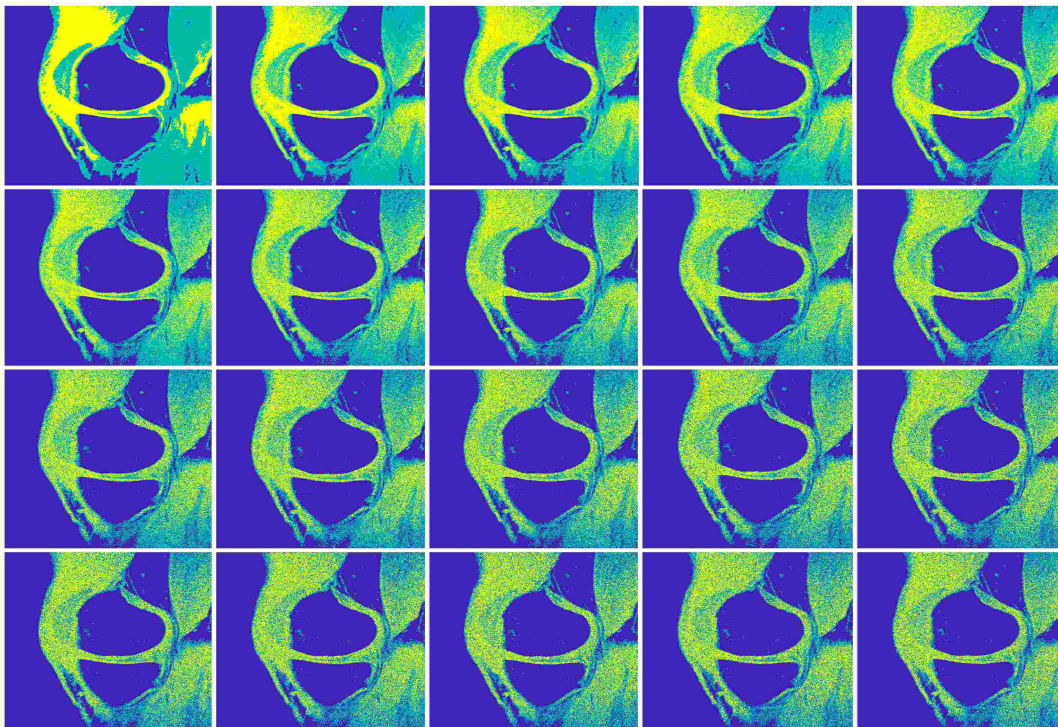
Obr.6: Fuzzy c-means pro 8 shluků implementována na obrazy se salt & pepper šumem, ($d=0 - 0,25$)



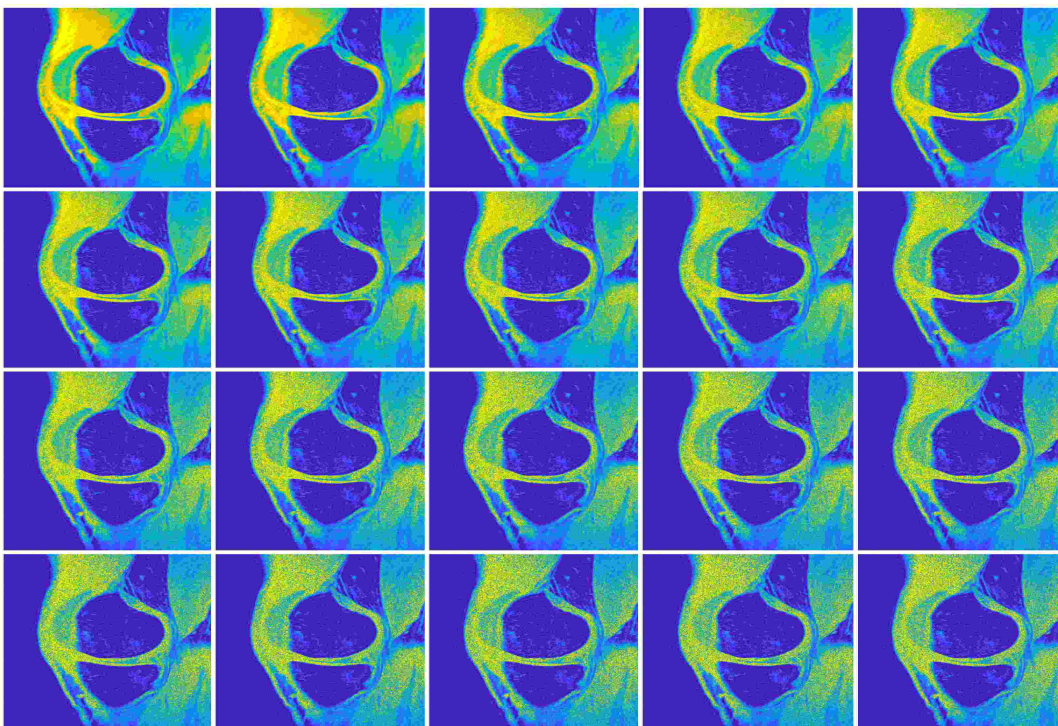
Obr. 1: K-means pro 3 shluky implementována na obrazy s gaussovským šumem, ($\mu=0$, $\sigma=0 - 0,25$)



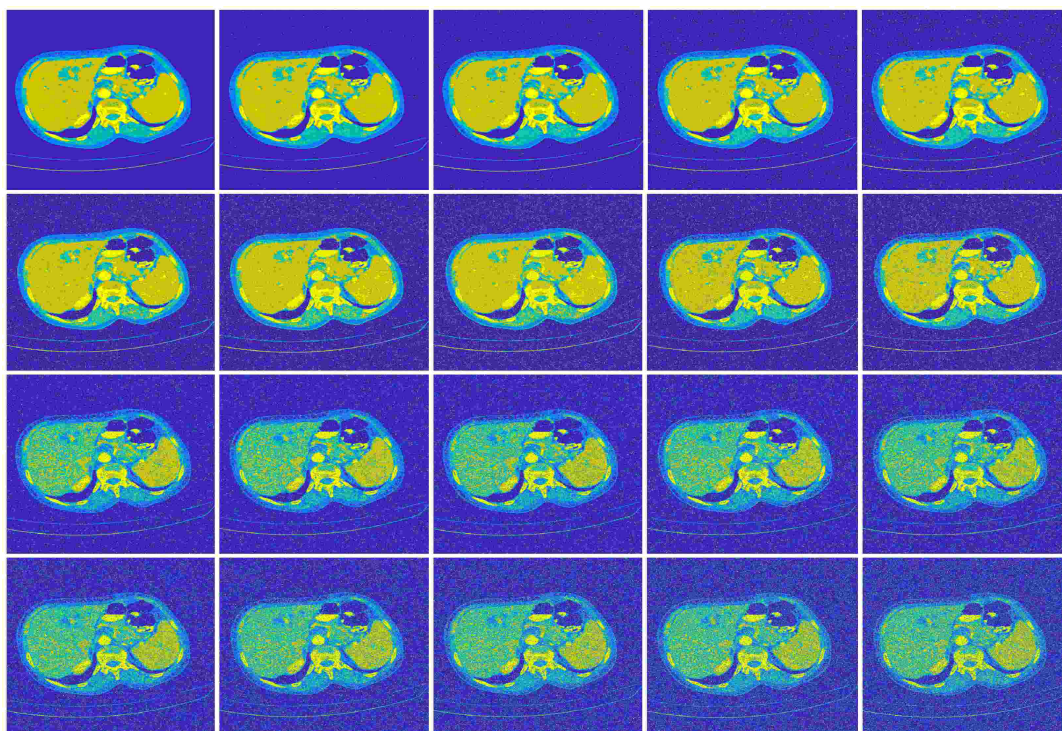
Obr. 2: K-means pro 8 shluků implementována na obrazy s gaussovským šumem, ($\mu=0$, $\sigma=0 - 0,25$)



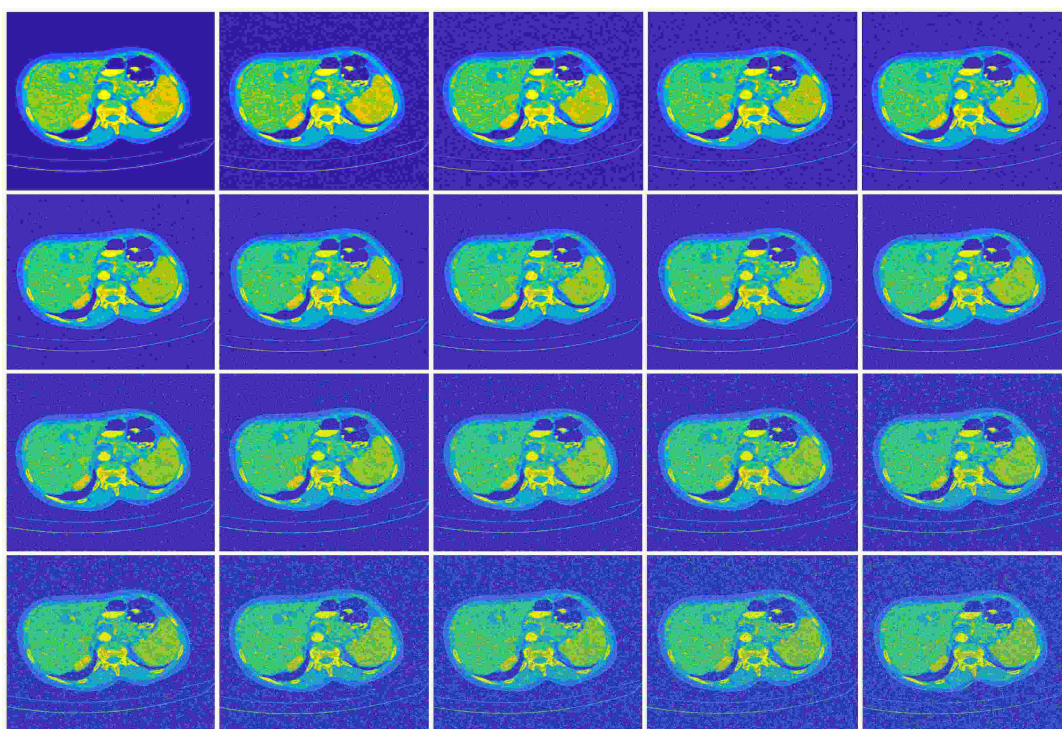
Obr. 3: Metoda k -means pro 3 shluky implementována na obrazy se speckle šumem, ($\mu=0$, $\sigma=0 - 0,25$)



Obr. 4: Fuzzy c -means pro 8 shluků implementována na obrazy se speckle šumem, ($\mu=0$, $\sigma=0 - 0,25$)



Obr. 5: Metoda fuzzy c-means pro 5 shluků implementována na obrázy se salt & pepper šumem, ($d=0 - 0,25$)



Obr. 6: Metoda fuzzy c-means pro 8 shluků implementována na obrázy se salt & pepper šumem, ($d=0 - 0,25$)

IV. Grafy a tabulky

Obr.1: Graf srovnání metod pro 3 shluky v závislosti na gaussovském šumu (vlevo), graf srovnání metod pro 3 shluky v závislosti na salt & pepper šumu (vpravo)

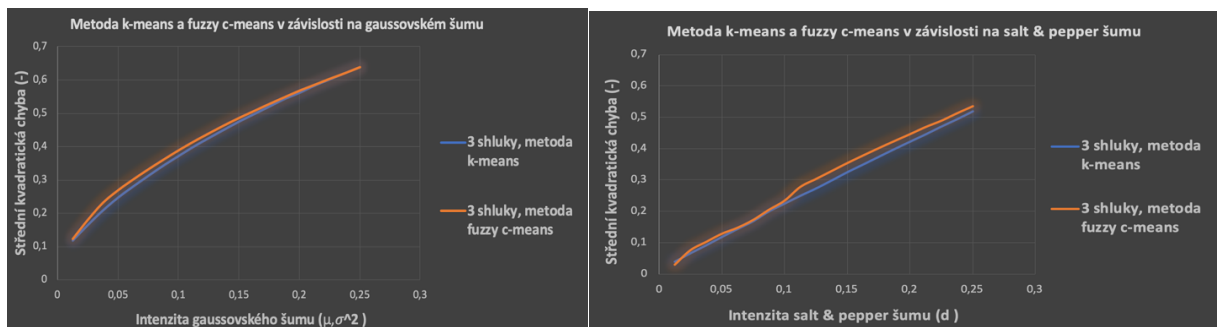
Obr.2: Graf srovnání metod pro 3 shluky v závislosti na speckle šumu (vlevo), graf srovnání metod pro 3 shluky v závislosti na localvar šumu (vpravo)

Obr.3: Míra ovlivnitelnosti jednotlivých datasetů šumem, vlevo speckle šum, vpravo localvar šum

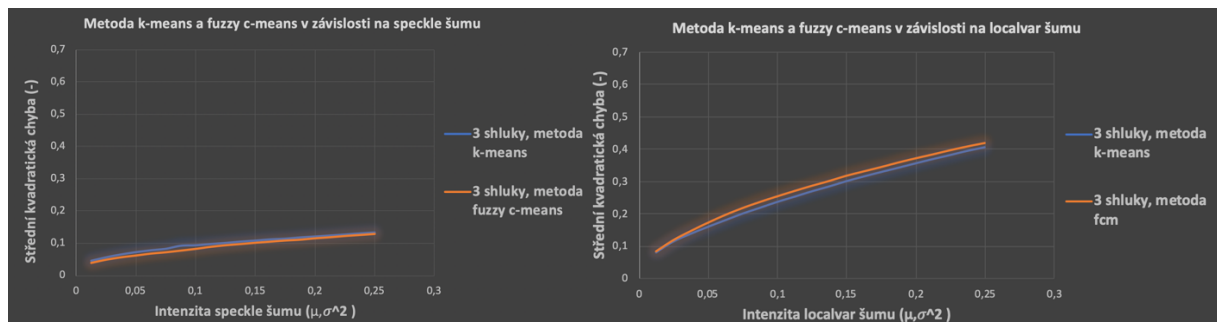
Obr.4: Míra ovlivnitelnosti jednotlivých datasetů salt & pepper šumem

Tab. 1: Výsledné hodnoty korelačních koeficientů

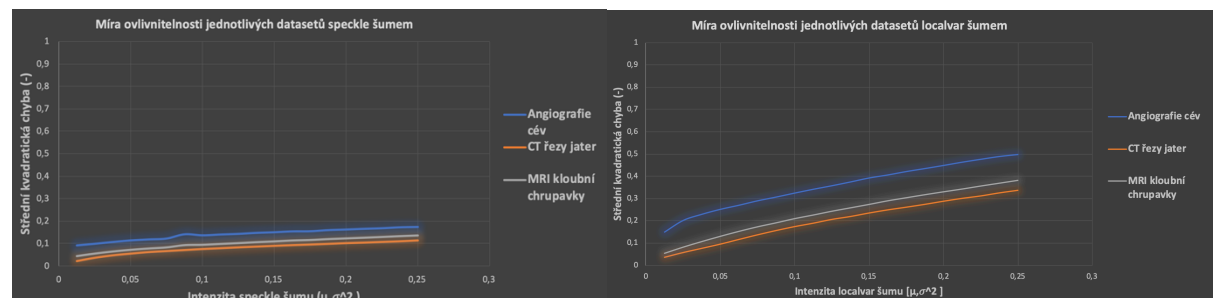
Tab. 2: Výsledné hodnoty střední kvadratické chyby



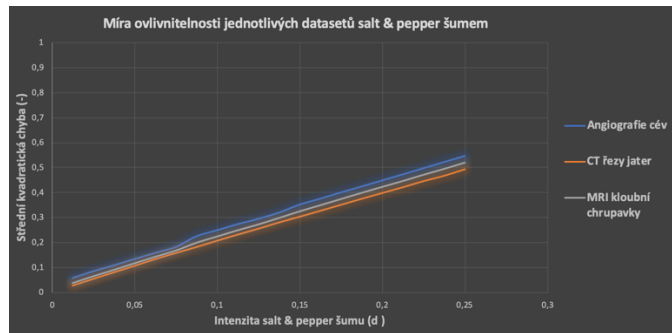
Obr. 1: Graf srovnání metod pro 3 shluky v závislosti na gaussovském šumu (vlevo), graf srovnání metod pro 3 shluky v závislosti na salt & pepper šumu (vpravo)



Obr. 2: Graf srovnání metod pro 3 shluky v závislosti na speckle šumu (vlevo), graf srovnání metod pro 3 shluky v závislosti na localvar šumu (vpravo)



Obr. 3: Míra ovlivnitelnosti jednotlivých datasetů šumem, vlevo speckle šum, vpravo localvar šum



Obr. 4: Míra ovlivnitelnosti jednotlivých datasetů salt & pepper šumem

Tab. 1: Výsledné hodnoty korelačních koeficientů

Hodnoty vytvořené na základě průměrů korelačních koeficientů 3 datasetů pro metodu k-means													
K-means, 3 shluky				K-means, 5 shluky				K-means, 8 shluky					
Intenzita šumu	Gaussovský šum	Salt & pepper šum	Speckle šum	Localvar šum	Gaussovský šum	Salt & pepper šum	Speckle šum	Localvar šum	Gaussovský šum	Salt & pepper šum	Speckle šum	Localvar šum	Gaussovský šum
1	0.847	0.944	0.959	0.893	0.883	0.961	0.978	0.924	0.905	0.963	0.983	0.933	
2	0.794	0.910	0.947	0.932	0.831	0.932	0.887	0.854	0.854	0.932	0.977	0.903	
3	0.750	0.881	0.936	0.824	0.791	0.905	0.962	0.860	0.815	0.904	0.972	0.878	
4	0.714	0.852	0.927	0.799	0.756	0.879	0.957	0.837	0.781	0.880	0.967	0.856	
5	0.684	0.826	0.920	0.778	0.727	0.852	0.952	0.816	0.751	0.856	0.962	0.836	
6	0.658	0.802	0.914	0.758	0.700	0.829	0.947	0.797	0.724	0.834	0.958	0.818	
7	0.633	0.775	0.893	0.740	0.676	0.806	0.943	0.780	0.699	0.813	0.953	0.801	
8	0.611	0.753	0.889	0.724	0.654	0.778	0.939	0.765	0.677	0.793	0.950	0.786	
9	0.591	0.733	0.884	0.709	0.635	0.758	0.935	0.750	0.656	0.773	0.945	0.771	
10	0.572	0.713	0.879	0.694	0.616	0.740	0.931	0.736	0.637	0.754	0.942	0.757	
11	0.554	0.693	0.874	0.681	0.599	0.721	0.928	0.723	0.619	0.737	0.938	0.744	
12	0.538	0.676	0.869	0.668	0.583	0.700	0.924	0.711	0.602	0.715	0.934	0.731	
13	0.522	0.658	0.865	0.656	0.568	0.683	0.920	0.700	0.587	0.698	0.930	0.720	
14	0.508	0.641	0.861	0.645	0.553	0.664	0.917	0.688	0.572	0.682	0.927	0.709	
15	0.495	0.624	0.857	0.634	0.540	0.648	0.912	0.678	0.559	0.665	0.923	0.699	
16	0.482	0.624	0.852	0.624	0.528	0.632	0.908	0.668	0.546	0.668	0.919	0.688	
17	0.470	0.592	0.848	0.613	0.516	0.617	0.904	0.659	0.534	0.635	0.916	0.679	
18	0.459	0.576	0.845	0.604	0.505	0.599	0.901	0.649	0.522	0.619	0.913	0.669	
19	0.449	0.561	0.840	0.595	0.494	0.585	0.897	0.641	0.512	0.605	0.909	0.661	
20	0.439	0.545	0.836	0.586	0.484	0.570	0.894	0.632	0.502	0.609	0.903	0.652	
Celkový průměr:	0.589	0.718	0.885	0.704	0.640	0.743	0.931	0.745	0.653	0.756	0.941	0.764	
Hodnoty vytvořené na základě průměrů korelačních koeficientů 3 datasetů pro metodu fuzzy c-means													
Fuzzy c-means, 3 shluky				Fuzzy c-means, 5 shluky				Fuzzy c-means, 8 shluky					
Intenzita šumu	Gaussovský šum	Salt & pepper šum	Speckle šum	Localvar šum	Gaussovský šum	Salt & pepper šum	Speckle šum	Localvar šum	Gaussovský šum	Salt & pepper šum	Speckle šum	Localvar šum	Gaussovský šum
1	0.847	0.964	0.964	0.892	0.873	0.964	0.975	0.918	0.891	0.964	0.985	0.930	
2	0.784	0.921	0.953	0.851	0.821	0.933	0.968	0.879	0.838	0.933	0.979	0.895	
3	0.737	0.892	0.943	0.819	0.779	0.904	0.963	0.851	0.797	0.908	0.974	0.868	
4	0.701	0.864	0.936	0.791	0.745	0.879	0.958	0.827	0.762	0.884	0.969	0.844	
5	0.672	0.840	0.929	0.768	0.716	0.854	0.954	0.807	0.733	0.859	0.965	0.823	
6	0.646	0.815	0.922	0.747	0.690	0.827	0.949	0.788	0.706	0.837	0.961	0.804	
7	0.623	0.780	0.915	0.729	0.666	0.805	0.945	0.771	0.682	0.816	0.958	0.787	
8	0.601	0.756	0.907	0.713	0.645	0.785	0.941	0.755	0.660	0.796	0.954	0.771	
9	0.581	0.724	0.900	0.698	0.626	0.767	0.938	0.741	0.640	0.776	0.950	0.757	
10	0.563	0.705	0.893	0.684	0.607	0.748	0.934	0.727	0.621	0.755	0.947	0.743	
11	0.546	0.685	0.889	0.671	0.590	0.727	0.931	0.714	0.604	0.737	0.944	0.730	
12	0.530	0.663	0.884	0.658	0.575	0.710	0.928	0.702	0.588	0.720	0.940	0.718	
13	0.516	0.646	0.879	0.646	0.559	0.694	0.924	0.691	0.573	0.703	0.937	0.706	
14	0.502	0.629	0.875	0.635	0.546	0.678	0.921	0.680	0.559	0.686	0.934	0.695	
15	0.489	0.614	0.872	0.625	0.532	0.662	0.918	0.669	0.546	0.670	0.931	0.684	
16	0.476	0.614	0.867	0.614	0.520	0.647	0.914	0.660	0.533	0.654	0.927	0.675	
17	0.465	0.585	0.863	0.605	0.509	0.633	0.911	0.650	0.522	0.639	0.925	0.665	
18	0.454	0.572	0.858	0.596	0.498	0.617	0.908	0.641	0.511	0.624	0.921	0.656	
19	0.444	0.566	0.857	0.587	0.487	0.603	0.905	0.633	0.500	0.610	0.919	0.647	
20	0.434	0.543	0.850	0.579	0.478	0.589	0.901	0.625	0.490	0.596	0.915	0.638	
Celkový průměr:	0.581	0.718	0.898	0.695	0.623	0.751	0.934	0.736	0.638	0.758	0.947	0.752	

Tab. 2: Výsledné hodnoty střední kvadratické chyby

Hodnoty vytvořené na základě průměrů střední kvadratické chyby 3 datasetů pro metodu k-means													
K-means, 3 shluky													
K-means, 3 shluky				K-means, 5 shluky				K-means, 8 shluky					
Intenzita šumu	Gaussovský šum	Salt & pepper šum	Speckle šum	Localvar šum	Gaussovský šum	Salt & pepper šum	Speckle šum	Localvar šum	Gaussovský šum	Salt & pepper šum	Speckle šum	Localvar šum	Localvar šum
1	0.118	0.038	0.046	0.080	0.279	0.082	0.057	0.166	0.625	0.182	0.098	0.382	
2	0.165	0.065	0.057	0.114	0.420	0.147	0.074	0.264	1.003	0.349	0.131	0.596	
3	0.209	0.091	0.066	0.138	0.544	0.210	0.094	0.340	1.321	0.504	0.161	0.787	
4	0.247	0.118	0.073	0.159	0.658	0.269	0.108	0.402	1.613	0.638	0.192	0.968	
5	0.281	0.143	0.078	0.180	0.756	0.339	0.118	0.464	1.863	0.782	0.221	1.125	
6	0.313	0.168	0.083	0.168	0.850	0.403	0.129	0.521	2.118	0.919	0.247	1.264	
7	0.343	0.199	0.093	0.218	0.934	0.473	0.139	0.578	2.340	1.059	0.274	1.407	
8	0.372	0.224	0.094	0.237	1.014	0.536	0.148	0.626	2.559	1.199	0.298	1.544	
9	0.400	0.249	0.098	0.253	1.089	0.600	0.159	0.676	2.775	1.341	0.328	1.671	
10	0.426	0.273	0.102	0.269	1.157	0.663	0.168	0.718	2.969	1.487	0.346	1.797	
11	0.451	0.298	0.106	0.285	1.222	0.724	0.176	0.764	3.190	1.630	0.373	1.909	
12	0.475	0.325	0.109	0.301	1.287	0.785	0.185	0.807	3.367	1.790	0.399	2.045	
13	0.498	0.349	0.113	0.315	1.349	0.846	0.193	0.846	3.516	1.938	0.419	2.134	
14	0.520	0.374	0.115	0.329	1.415	0.905	0.202	0.887	3.700	2.090	0.443	2.252	
15	0.543	0.399	0.119	0.343	1.458	0.967	0.215	0.923	3.848	2.245	0.467	2.341	
16	0.562	0.422	0.122	0.366	1.512	1.028	0.226	0.957	3.998	2.399	0.489	2.441	
17	0.583	0.446	0.125	0.369	1.565	1.090	0.234	0.992	4.159	2.549	0.511	2.540	
18	0.602	0.471	0.128	0.382	1.608	1.151	0.242	1.028	4.284	2.697	0.532	2.643	
19	0.621	0.494	0.132	0.395	1.657	1.207	0.250	1.059	4.422	2.856	0.558	2.736	
20	0.639	0.519	0.135	0.406	1.693	1.270	0.259	1.091	4.517	3.014	0.580	2.831	
Celkový průměr:	0.418	0.283	0.100	0.266	1.123	0.685	0.169	0.705	2.909	1.583	0.353	1.771	
Hodnoty vytvořené na základě průměrů střední kvadratické chyby 3 datasetů pro metodu fuzzy c-means													
Fuzzy c-means, 3 shluky				Fuzzy c-means, 5 shluky				K-means, 8 shluky					
Intenzita šumu	Gaussovský šum	Salt & pepper šum	Speckle šum	Localvar šum	Gaussovský šum	Salt & pepper šum	Speckle šum	Localvar šum	Gaussovský šum	Salt & pepper šum	Speckle šum	Localvar šum	Localvar šum
1	0.123	0.029	0.039	0.085	0.339	0.082	0.066	0.210	0.905	0.231	0.106	0.565	
2	0.180	0.075	0.049	0.119	0.502	0.156	0.085	0.319	1.374	0.424	0.140	0.859	
3	0.231	0.102	0.056	0.147	0.638	0.234	0.098	0.399	1.750	0.585	0.174	1.091	
4	0.269	0.128	0.062	0.173	0.754	0.317	0.109	0.476	2.081	0.747	0.206	1.306	
5	0.301	0.147	0.068	0.197	0.857	0.389	0.120	0.540	2.382	0.913	0.235	1.490	
6	0.332	0.172	0.072	0.218	0.949	0.465	0.132	0.602	2.668	1.074	0.261	1.663	
7	0.361	0.204	0.077	0.237	1.036	0.535	0.141	0.658	2.915	1.216	0.287	1.824	
8	0.389	0.234	0.082	0.254	1.117	0.597	0.151	0.711	3.159	1.358	0.311	1.980	
9	0.415	0.277	0.089	0.271	1.194	0.654	0.160	0.761	3.387	1.512	0.336	2.125	
10	0.440	0.303	0.094	0.287	1.268	0.717	0.169	0.809	3.592	1.667	0.360	2.260	
11	0.464	0.328	0.097	0.302	1.340	0.774	0.178	0.855	3.793	1.779	0.382	2.395	
12	0.486	0.354	0.101	0.318	1.404	0.825	0.186	0.898	3.981	1.898	0.406	2.517	
13	0.507	0.378	0.105	0.331	1.468	0.856	0.194	0.939	4.175	2.043	0.432	2.638	
14	0.528	0.401	0.108	0.344	1.528	0.908	0.203	0.979	4.339	2.186	0.455	2.757	
15	0.548	0.424	0.111	0.359	1.585	0.967	0.211	1.020	4.507	2.340	0.477	2.866	
16	0.568	0.446	0.115	0.371	1.639	1.027	0.219	1.058	4.661	2.482	0.502	2.978	
17	0.586	0.469	0.118	0.384	1.688	1.084	0.227	1.092	4.805	2.629	0.524	3.081	
18	0.604	0.489	0.122	0.396	1.736	1.124	0.235	1.128	4.947	2.781	0.545	3.185	
19	0.621	0.513	0.126	0.408	1.783	1.180	0.244	1.164	5.093	2.926	0.567	3.285	
20	0.639	0.535	0.129	0.419	1.829	1.238	0.255	1.194	5.229	3.077	0.589	3.384	
Celkový průměr:	0.430	0.300	0.091	0.281	1.233	0.706	0.169	0.791	3.487	1.693	0.365	2.212	

V. Příloha na CD

Název	Popis
Kod a Algoritmy	Praktická část práce v programu MATLAB
2019_SKA0143_BP	Bakalářská práce v elektronické podobě
Data	Vstupní obrazová data, křivky EKG, výsledné tabulky hodnot